High School General Relativity

(draft)

© 2005 - 2012 Steven J. Heilig

sheilig@spa.eduIntroduction: Why Study General Relativity?

There are really two theories at the heart of physics today: Quantum Mechanics and General Relativity. (Quantum Mechanics deals with phenomena that largely appear in the study of the very small: atoms, molecules, photons, and so on. In Quantum Mechanics many variables are found to have discrete values, in contrast to classical physics – General Relativity included – in which variables like energy can take on a continuous range of values.) Everything that we claim to understand in physics is ultimately based on one of these two theories. To know physics, you need to spend some time with each of these. They provide the post-Newtonian insights that have completely changed our ideas of time, space, energy, mass, and everything else you can think of. My hope is to give you the basis to understand the fundamentals of the field, to open the door for you. You may choose to step through.

If you are interested in philosophy, these ideas of how the universe works are very provoking. If you want to understand how the world really works, these ideas are critical. If you love being confronted with challenging ideas about the fundamental nature of the universe, these topics are fascinating.

The following work represents an effort to provide the needed text support to enable you to come to some understanding of General Relativity. Try it out. You’ll grab your head in fits of brain overload. You’ll laugh, and perhaps cry, but you will definitely confront big questions. Don’t be afraid of those; whoever figures out satisfying answers will likely earn a Nobel Prize. Have a ball.

Chapter 1. The Postulates of General Relativity

Special Relativity is a theory designed for observers in uniform motion relative to each other (a *special* case of relative motion). A more general theory would allow us to predict, for example, what O’Malley (a moving observer, O’) would observe based on Ophelia’s measurements (Ophelia is the “at rest” observer, O) even if O’Malley were accelerating. Einstein set about constructing (discovering?) such a theory immediately upon completion of Special Relativity, knowing that the Special theory was not complete. To get to Einstein’s basis for this more general theory, we need to go back to Newton’s mechanics. Einstein began his work with two postulates: The first is called the equivalence principle, and is the major focus of this chapter; the second simply states that in the limit of no acceleration between observers, the results of any new theory should reduce to those of Special Relativity.

**Inertial vs. Gravitational Mass**

If you want to examine the motion of an electrically charged object in an electric field, you would use Newton’s 2nd law, F = ma. The mass, m, represents an object’s ability to resist changes in motion, otherwise known as inertia. The force, F, is proportional to the amount of charge on the object. So there are two properties of this object that determine its motion in an electric field: its inertia and its charge.

A similar relation holds for the motion of an object in a gravitational field. The mass of an object comes into the right-hand side of Newton’s 2nd law (F = m∙a) as the inertia of the object. What determines the left-hand side, the gravitational force? In Newton’s Law of Universal Gravitation, mass plays a role analogous to electric charge in the example above: the strength of the gravitational force is proportional to the mass. These are two different functions for mass – determining inertia and determining the strength of the force – and they are called inertial mass and gravitational mass, respectively. Just as the electric charge and inertia for a charged object are not dependent on each other, there is no *a priori* reason why inertial mass and gravitational mass need to be the same.

Newton understood this distinction. What does this imply about the acceleration of various objects dropped in a gravitational field? We normally calculate the acceleration of gravity (for an object of mass m) near the surface of the Earth as follows:

Begin with Newton’s 2nd Law, F = m∙a. Suppose that force is the force of gravity, F = . Putting these together, we find

 = m∙a,

which simplifies (by dividing out the mass of the object) to:

acceleration due to gravity = . Done.

Note that the mass of the object has canceled out – all objects should have the same acceleration. But now suppose we are explicit that the inertia of an object could be different from the gravitational mass. We will use mi to represent the inertial mass and mg to represent the gravitational mass.

 = mi∙a

now becomes:

acceleration due to gravity = .

The acceleration due to gravity is now seen to depend on the ratio of the two types of mass. If two objects had different ratios, they would accelerate at different rates, and Galileo’s conclusion that all objects accelerate equally under the influence of gravity would be false. One could simply drop balls made of different materials, say carbon and lead, and see if one or the other always lands first. In simple demonstrations one can see that different objects seem to accelerate equally, at least roughly. But there are other ways to achieve higher precision. Newton himself built pendula of different materials to test this idea more precisely. Again, the driving force is gravity, while the inertia of the pendulum bob opposes the acceleration. He measured the period of the pendula to see if there was a difference. Sir Isaac determined that the periods, and hence the accelerations, were the same to within an uncertainty of one part in one thousand. In the late 1800s, Ernst Mach and Roland von Eötvös built a torsional pendulum which responded slightly to the Earth’s rotation and found the masses to be equivalent to 5 parts in one billion. (Experiments using torsional pendula for measuring small forces are now generally referred to as Eötvös experiments.) Robert Dicke in the 1960s got the uncertainty down to one part in 100 billion in comparing aluminum to copper. And by looking at how the Sun might pull differently on the Earth and the Moon (the Earth has a higher proportion of heavy elements) the equivalence principle will be put to the test in an experiment called APOLLO (Apache Point Observatory Lunar Laser-ranging Operation) which is currently taking data. It is hoped that enough data will have been collected in the next year to provide a stringent test of mass equivalence on a large scale. So far, the ratio of inertial and gravitational masses seems to be the same for all materials.

**The First Postulate: The Equivalence Principle**

Einstein took this equality of masses to be no mere coincidence. But why should they be equal? Consider the following situation: If you are in deep space and release two objects inside your rocket ship, they will just float there together. If you now fire your rocket engine, they will hit the “floor” at the same time simply because they remain in (identical) uniform motion and the floor comes up to meet them. And as we have seen, gravity causes the same result. If the effects of gravity were the same as the effects of an accelerating frame of reference, that would explain why the two types of masses would have to come out the same.

Einstein took this as his first postulate of General Relativity, called the Equivalence Principle:

The local effects of gravity are equivalent to the effects of an accelerating frame of reference.

**Implications of the First Postulate**

The first implication of this postulate is, as described above, any two objects should accelerate toward the floor at the same rate; this has been tested with good precision and found to be within uncertainty.

In order to figure out further implications, we need a trustworthy observer. In our study of Special Relativity, Ophelia played the role of the observer at rest relative to the Earth. We could always trust her observations, since she was clearly an inertial observer – she was not accelerating, she wasn’t even moving. For the case of General Relativity, just who is Ophelia supposed to be? You might say an observer standing still on the Earth, but before you bet your college savings on that consider Einstein’s first postulate. If gravity is equivalent to an acceleration, what does that say about an observer standing on the Earth? That observer would certainly feel gravity. Therefore that observer would see things the same way an observer accelerating at 9.8 m/s2 out in deep space would. So the old Ophelia is equivalent to an accelerating observer, and is not inertial.

**Fictitious forces**

Back in the study of Newton’s Laws, we discussed fictitious forces, the centrifugal and Coriolis forces in particular. Let’s consider a simpler case right now. Think about a room just like the room you’re sitting in now. Imagine it to be in deep space, far away from the Earth or any other source of gravity. Everything is floating around you. But underneath the floor is a rocket engine. When it fires, it accelerates the room in the upward direction at 9.8 m/s2. If you are in the room and let go of a rubber ball, it will remain in uniform motion (constant velocity) while the room accelerates upward. What will it look like? Relative to the spaceship it will fall to the floor, appearing to accelerate downward as it goes. Just as if there were gravity. So it appears to you that there is a downward force, which is in fact fictitious, that accelerates the ball at 9.8 m/s2.

If you want to hold the ball at rest in your hand, you need to apply a force strong enough to accelerate the ball at the same rate the room is accelerating. To see how much force you’d need, just use Newton’s 2nd Law of Motion: F = m∙a. The force needed to hold the ball “up” is just the mass times the acceleration. So the fictitious downward force, in fact any fictitious force, is proportional to the mass of the object. And gravity is proportional to mass. Coincidence? Einstein thought not.

Please keep in mind the directions: if your room is accelerating up, toward the ceiling, that’s the same as gravity down; if you’re in a car that accelerates forward, it feels like a backward force pushing you into your seat.

**Back to Ophelia**

So where do we go to find a truly inertial observer? The easy answer is an observer far away from any large masses, perhaps in an empty region of space between galaxy superclusters. Easy to say, but not very practical.

But there is another answer that works on the surface of the Earth. Since acceleration up equals gravity down, acceleration down would equal gravity up. If Ophelia were accelerating downward at 9.8 m/s2 she would experience an upward (fictitious) gravity with that magnitude. The net force on her would be the sum of the real gravity down plus the fictitious force upward, which is zero! Accelerating downward can cancel the effects of a gravitational field. To cancel the gravity on the surface of the Earth, then, Ophelia only needs to accelerate downward at 9.8 m/s2. And to do that, she only needs to jump off a cliff. (Don’t try this at home.) But the principle is that anyone or anything in freefall constitutes an inertial observer, and an observer in freefall will experience no gravity. Remember this; it will come in handy in the monkey-hunter problem below. In 1908, Einstein even talked to a man who had fallen off a roof in order to see what the man had experienced on the way down!

Here’s the kicker: Will Ophelia, in freefall, experience any gravity? From what we’ve said, the answer is no. If she holds a pen in front of her and lets go, it will remain at rest *relative to her*. So our most trusted observer, an inertial observer, says there is no gravity! Therefore gravity only shows up due to our frame of reference, and it is proportional to mass. It begins to sound like gravity, in some way, is a fictitious force. More on this later.

**The Monkey-Hunter problem**

Back in the unit on Newton’s Laws, we studied projectile motion. One problem you saw demonstrated was called the Monkey-Hunter problem. The hunter aims directly at the monkey; at the instant the hunter shoots the monkey drops from the branch. As we found, the bullet and the monkey fall the same amount from where they would have been without gravity and the monkey gets hit. (And it didn’t matter how fast the bullet was going as long as it made it over to the monkey before the bullet and the monkey hit the ground.) It’s an amazing thing.

How can the first postulate help us decide if the monkey will get hit? Let’s imagine an observer (Ophelia) in freefall. What will happen? In Ophelia’s view there is no gravity. The rifle is aimed at the monkey. The bullet, shot by O’Malley and aimed at the monkey, follows a straight path. The monkey, who doesn’t accelerate downward relative to Ophelia, has to get hit.

You might object that the bullet follows a curved trajectory because of gravity – we can’t just throw gravity out. But Ophelia, the monkey, and the bullet are accelerating at the same rate in the same direction (by O’Malley’s observations). They are not accelerating *relative to each other*. If Ophelia maps out the trajectory of the bullet in her frame of reference, she will observe straight line motion, following Newton’s 1st Law. So, as stated above, the monkey gets hit. And although different observers can disagree about a lot of things, they cannot disagree about whether an event happens. If the monkey gets hit, the monkey gets hit. So we’ve solved the problem without the use of any equations. Very cool. And no real monkeys were harmed in the performance of this gedanken (thought) experiment.

**Tidal Effects**

There is one word in Einstein’s first postulate that has not come up so far: “local.” The *local* effects of gravity are equivalent to the effects of an accelerating frame of reference. Why did we need that word?

Suppose we imagine things happening that are not local, like Ophelia and O’Malley dropping balls in different places. If those two places are in different parts of a room that is accelerating upward, both balls will fall straight down. But suppose they are in St. Paul and Paris. They may both fall straight down, but now “down” has another meaning: toward the center of the Earth. The balls do not follow parallel paths.

Another case is if Ophelia and O’Malley are at different elevations. Since gravity depends on the distance from the center of the Earth, gravity will cause one ball to accelerate faster than the other. If they each measure the acceleration of their ball very carefully, they will get slightly different values. In the case of the accelerating room, the acceleration is of the same magnitude everywhere.

So there are a couple of ways we can tell the difference between the gravitational field of a star or planet and an acceleration, but they both involve comparing things over a wide area. Any *local* test will not distinguish between a gravitational field and an acceleration. That’s the Equivalence Principle.

These effects over large regions are called “tidal effects.” They have been known of since the time of Newton, who first explained why the Earth has tides: We have ocean tides on the Earth because the strength of the gravitational pull by the Moon is stronger on one side of the Earth and weaker on the other, so we get two high tides-one roughly oriented toward the Moon and the other opposite the first. That’s how the name came about. People who work in the field of General Relativity say that tidal effects are the only thing real about gravity. And the closer you are to a source of gravity, the bigger the tidal effects will be.

We will save a detailed calculation of tidal forces until the chapter on black holes. For us on the Earth, if we were to jump off a chair, how big would the tidal force be that is trying to pull our head and feet apart? The result is less than one-tenth of a milliNewton. Wow. No wonder we haven’t talked about it before. This is a tiny effect for us personally. But the Earth is pretty big, so the Moon can raise noticeable ocean tides.

**The Second Postulate**

This chapter has focused on the equivalence principle, but there was a second postulate as well.

Einstein’s second postulate:

In the limiting case of zero acceleration, the results of Special Relativity should apply. (In other words, don’t mess with Special Relativity!)

So the basic idea is that if you take any of the formulas of General Relativity and plug in a zero for the acceleration, you should wind up with the corresponding formula for Special Relativity. But also, within that second postulate, there is an important idea that should not be forgotten. One of the postulates of Special Relativity said that the laws of physics should be the same for all observers. (Galileo used this idea in his version of relativity. He believed in the Copernican (Sun-centered) model of the solar system, and needed to justify how the Earth could be in motion around the Sun and we wouldn’t feel the effects of this motion.) Einstein said the goal for General Relativity was to discover the laws of gravity and motion that would be the same for all observers. This gets more complicated than before, because now we must accept the possibility of accelerated frames of reference. This principle is called “general covariance.”

**A final word**

Once upon a time, Ophelia fell off the top of a very tall building. On the way down she passed O’Malley on a balcony. He shouted out, “How’s it going?” She replied, “Fine so far!”**Exercises**

1.1 It was said that to an observer in freefall there is no gravity. Let’s check out one case. Suppose at some time Ophelia is in freefall, and she is traveling downward at 3 m/s. At this time, a ball is launched upwards at 15 m/s.  
  
a) What is the speed of the ball relative to Ophelia?  
  
b) 1 second later, what are the speeds of Ophelia, the ball, and the ball relative to   
 Ophelia?  
  
c) What will the speed of the ball be relative to Ophelia for other times?

1.2

Imagine a group of four marbles spread out on an imaginary circle as shown. If they are released from rest and allowed to fall freely, how will they move relative to each other? (Consider only the force of gravity from the planet acting on each marble; ignore their mutual attraction.)

1.3 Design a device using springs and masses to measure tidal effects inside your spaceship. (It could be used as an alarm in case you get too close to a black hole.)

1.4 What will remain of gravity if you fall freely in a gravitational field?

1.5 Does general covariance mean that the forces you feel will be the same in all frames of reference?

1.6 You get the opportunity to solve a projectile problem in the ‘no gravity’ (freefall) frame of reference. Imagine a projectile being launched from the floor (yi = 0) with a speed of 5 m/s at an angle of 60°. If you are falling freely, it will follow a straight line path relative to you. As a result, its equations of motion are  
  
x = vxt and y = vyt, where vx = 5cos 60° and vy = 5sin 60°.  
  
The floor is at the same height as the ball initially (yi floor = 0), but the floor is accelerating upward. (Remember, gravity down is the same as accelerating the room upward.) So the equation of motion for the floor is  
  
yfloor = (½)gt2.  
  
a) At what time will the ball and the floor be at the same y coordinate?  
  
b) What will the x coordinate of the ball be at that time?  
  
You have just calculated where the ball will hit the ground. Congratulations!

1.7 Suppose you are in your car and you accelerate forward. Your fuzzy dice hang backward at an angle of 20°. What is your acceleration? (Hint: Think like Einstein.)

1.8 Why do fictitious forces appear for non-inertial observers?

Chapter 2. Non-Euclidean Geometry

**A fable**

One day Ophelia and O’Malley woke to find themselves on different worlds. They had walkie-talkies that allowed them to communicate with each other, but they had no idea where they were.

“What should we do?”

“I suppose we’d best look around.”

So they began to explore their surroundings. Each was in a forest, but it was very quiet. They each picked a direction and headed out. After a while, O’Malley said, “I must have been walking in a circle – I’m right back where I started.”

Ophelia responded, “Well, I’m not back where I started. I don’t know where I am just yet. Can you try again and be careful to go in a straight line?”

“Yah, I’ve got my laser pointer. I can use it to mark out a straight line. We’ll see what happens.”

Time passed as Ophelia continued on her path, too. After a time, O’Malley spoke again. “I’m back again. I mean, not just talking to you, but I’m back where I started. And I’m sure I didn’t turn so much that I would have completed a circle so quickly.”

Ophelia replied, “Well, I still am not back to my starting point, and I’ve been careful to do just as you have to be sure I’ve been moving in a straight line. But you know, I’ve come across the most odd statue here. It’s a seated figure, looks like it’s puzzled. How are we going to figure this place out? Eeek! It’s starting to move!”

The large stone figure arose, and spoke in a deep, gravelly voice. “Try measuring the sum of the angles in a large triangle.” And then it sat down again, as still as a rock, which of course it was.

“Did you hear that?” Ophelia asked.

“Yes, but I don’t know what that’s got to do with anything. Was the statue talking to you or me? I’ll give it a try, though. I’m stumped,” answered O’Malley.

“I’ll give it a try, too,” said Ophelia.

They agreed to pace off equilateral triangles; each would make a 60° angle relative to their original path and travel for 5 minutes. That went well. After that period, they made another 60° angle and resumed their travels. But before the next 5 minutes was up O’Malley called Ophelia. “Hey, I’ve just come to the path I was following originally. A bit earlier than expected, I suppose. And the weirdest part is that the angle I get here isn’t 60°, it’s bigger, more like 70°.”

“Well, I’m not back to my original line, so I’ll keep going. Maybe I’m slowing down,” she replied. After another few minutes, she finally came across her original path. “O’Malley, old chap, guess what? I’ve found my original path, too, and the angle I get here is more like 45°. I’m going to retrace my original path back to the statue to check on the length of this side of the triangle, you know, to see if I’ve been slowing down. It took longer than I expected to reconnect with my original path.”

“I’ll do the same,” said O’Malley. Shortly, he spoke up again, “I’m back to the beginning of my triangle, again in less than 5 minutes. So I’ve got a ‘triangle’ with two short sides, one longer side, and angles of 60°, 60°, and 70°.”

By the time Ophelia arrived back at the statue, she found that this side of the triangle was also over 5 minutes in travel time, and equal to the second side but longer than the first side that began at the statue. “I don’t get it,” Ophelia said. “I’ve got a triangle with two 60° angles and one 45°, and two sides equal. What’s going on?”

Again, the stone statue rose and after a small pause it spoke, “You are both walking on curved surfaces. Your geometry is not Euclidean.” And then it sat down again.

“Of course, it stands to reason!” cried Ophelia.

The End (of the fable)

**What’s this got to do with General Relativity?**

You may well be asking this question. The answer is coming, but before we get to it there is another puzzle that has a related answer. This puzzle grows out of Special Relativity and was asked by a student:

Suppose you are in a spaceship traveling at a significant fraction of the speed of light. But you’re going in a circle. From Special Relativity we know that distances in the direction of travel are contracted, but distances perpendicular to the motion are unaffected. What does this say about the relationship between the radius and the circumference of the circle you’re following? After one time around, will your odometer read a value equal to 2r? (I’ll bet you didn’t know spaceships came with odometers! Well, ours do.) Length contraction will shorten the circumference you measure, but the radius, being perpendicular to the direction of motion, is unaffected.

How can we visualize the circumference not being equal to 2r? Consider the surface of a sphere. Let’s choose a point, call it A, and imagine a set of points equidistant from it. You could find these points by taking a short piece of string, anchor one end at A, and swing the other end around to construct a circle. Or use a flexible ruler and mark out a bunch of points and connect them. It will be just like a line of latitude on the Earth, everywhere the same distance from the pole. If I try to measure the radius of this circle on the sphere, I run into a problem. The center of this circle is not really on the sphere, but somewhere inside. But if I restrict myself to working on the surface, like an ant geometer would, the point A is the center. The radius is then the length of the piece of string I used to construct the circle. And this string followed the curve of the sphere as I swung it around. This measurable radius is longer than the ‘true’ radius inside the sphere. As a result, the circumference I find by multiplying 2 times the measured radius will be bigger than the actual circumference. So this is a case where the circumference will be less than 2r. The geometry of curved surfaces can produce unanticipated results.

Let’s get back to the spaceship. It also found a circle where the circumference was less than 2r. While it is true that the spaceship is accelerating (centripetally) so Special Relativity does not exactly apply, if the circle is big enough the acceleration can be pretty small and Special Relativity should be a reasonable approximation. So here we are, with a small acceleration and spatial measurements that mimic a curved surface. Einstein used his Equivalence Principle to say that if the effects of an acceleration can look like curved space, then the same will be true of gravity. In other words, he saw that it might be possible to describe gravity as an effect of curved spacetime rather than as a force acting at a distance between masses.

Before going into more of General Relativity, we will spend some time learning about non-Euclidean geometry. And by the end of this chapter, we will understand what is happening with O’Malley and Ophelia.

**Non-Euclidean geometry**

Non-Euclidean geometry is geometry that does not follow all of Euclid’s postulates.

Here are Euclid’s first four postulates:

1. It is possible to draw one and only one straight line from any point to any point.
2. From each end of a finite straight line it is possible to produce it continuously in a straight line by an amount greater than any assigned length.
3. It is possible to describe one and only one circle with any center and radius.
4. All right angles are equal to each other. (Although this 4th postulate may seem obvious, it is telling us is that right angles are the same no matter where you produce them – the plane is the same everywhere.)

The 5th postulate is somewhat more involved:

1. If a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

Here is a diagram that illustrates the 5th postulate:

A. This straight line falls on the other two straight lines

B. These two interior angles on the same side of the first line add up to less than 180°.

C. Therefore the two lines will meet (eventually) on this side.

This postulate is commonly known as the parallel line postulate. It is equivalent to saying that if you have a line and a point not on the line, there is only one line through the point which is parallel to the original line.

A. Original line

B. Point not on the line

C. Only one line goes through the point parallel to the original line

Right from the time of Euclid, who wrote his *Elements* around 300 B.C.E., geometers have not liked this postulate. It is much wordier than the prior postulates, and seems less fundamental. Much effort was given to prove it from the first four postulates, but this was never managed. In 1767, d’Alembert called the problem of the 5th postulate “the scandal of elementary geometry.” (www-groups.dcs.st-and.ac.uk, McTutor History of Mathematics, University of St. Andrews, Scotland) Finally, Saccheri decided to try a Reductio Ad Absurdum proof: He assumed a different version of the 5th postulate and hoped there would be a contradiction as a result. He developed a good deal of non-Euclidean geometry but could not accept the odd results. Soon after that Bolyai and Lobachevsky developed a more complete version of geometry in which there was more than one line through a point parallel to a given line (called hyperbolic geometry) and Riemann produced a geometry in which there were no lines through a point parallel to a given line (called elliptic geometry). And no contradictions arose. It was possible to do geometry (of a different type) with different versions of the 5th postulate. Non-Euclidean geometry was born.

We will learn a bit about geometries that follow the first four, but not the 5th postulate. To do so, we first need to know something about the basic unit for geometry, the line.

**Geodesics**

To do most types of geometry you need to be able to construct a straight line. The general name for a straight line, no matter what surface you’re looking at, is ‘geodesic.’ But what do we mean by a straight line on a curved surface? There are several definitions of straight lines on a plane: the shortest distance between two points, a set of points any two of which produce the same slope, a path which at any point has 180º on each side. On curved surfaces, we will define a straight line as the path connecting two points in the shortest (and sometimes longest) distance. These are our geodesics.

On a sphere, to find the shortest distance between two points you can take a piece of string and pull on the ends until it finds the shortest path on its own. Use this method on a globe to connect St. Paul and Athens and you’ll see why airline paths head north initially when Athens is south of St. Paul. (Note that if you head in exactly the opposite direction, you will again reach Athens, although this time your route will be the longest straight line distance.) Another method of constructing straight lines, when you don’t know both endpoints, is very simple: use a straightedge. But what type of straightedge will work on a curved surface? A flexible one. A ribbon or a strip of thin plastic will do. If you place a ribbon on a table and smooth it to the surface, it will follow a straight line. If the right edge won’t lie down, then the ribbon will show a path which bends to the right. But if you smooth it out, you’ll get a straight line every time. And this will be the shortest distance between two points: a geodesic. This works well on curved surfaces.

**Two Geometries**

For our purposes there are two commonly discussed types of non-Euclidean geometries: In Reimannian geometry, also called elliptical, the 5th postulate is changed to say that there are no parallel lines, which is true for surfaces of positive curvature, the simplest of which is a sphere; in the Lobachevskian case, also called hyperbolic, the 5th postulate says that there are infinitely many lines through a point not on a line which do not intersect that line, which applies to negatively curved surfaces such as a saddle. (Curvature will be discussed below.)

Let’s look at the surface of the sphere. If you take a ball and draw a straight line using a straightedge, you will find that you return to your starting point and begin to retrace your path. (Does this sound familiar?) Straight lines on a sphere are great circles, so called because they are the largest circle you can construct on a sphere. Lines of longitude are good examples, although lines of latitude are not (except for the equator). If you take one line of longitude, and a point on the equator not on that line, you might assume that the line of longitude through the point would be parallel to the first line. After all, the equator intersects both lines at right angles. But what happens to the two lines as you follow them either north or south? They intersect at the poles. Try as you might, you cannot have two lines that do not intersect (twice, in fact) on the surface of a sphere. In this sense there are no parallel lines on the surface of a sphere.

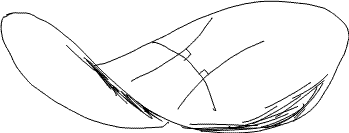
A

B

C

In addition to violating Euclid’s 5th postulate, you may have noticed something else. The triangle ABC has two right angles, one at A and one at B. What does that say about the sum of the angles in triangle ABC?

A straightforward example of negative curvature is the surface of a saddle. By using a straightedge you can make straight lines, although they can’t be characterized as easily as those on a sphere. But seen from above, they do not look straight. If you have two lines connected by a line segment which is perpendicular to both lines, the lines will curve away from each other as you follow them in either direction. If you start with a line and a point not on the line, you can then draw more than one line through that point that will never intersect the original line. There is an excess of parallel lines here. (That’s why it’s called hyperbolic.)



Since lines seem to curve away from each other in this geometry, what would a triangle look like? What does this suggest about the sum of the angles of a triangle in a world like this? (And whose trip does this remind you of?)

**The Curvature of a Surface**

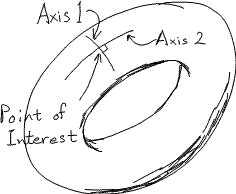
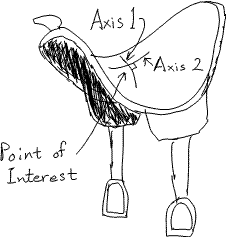
How to measure a radius of curvature

1) Imagine you have a one-dimensional curve in a plane, and picture a few radii projecting from it. Where they intersect is the center, and the length of each radius is the radius of curvature. You don’t need a complete circle to do this: a small arc is good enough.

r1

r2

2) For a curved surface in three dimensions, there are two radii you need to find. You begin by choosing the point at which you want to know the curvature. At that point, draw in two perpendicular axes. If the surface has some natural lines of symmetry, use those directions. Then find the radius of curvature for each axis at the point in question.

It is now straightforward to find the curvature of the surface from the two radii:



If the two centers are on the same side of the surface, you have positive curvature. In the case of the saddle, one center of curvature is above the saddle and the other is below – they are on opposite sides, so one radius is positive and the other negative: the product will be negative, and so will the curvature.

**The Characteristics of Positively and Negatively Curved Surfaces**

In addition to what has been said about parallel lines, there are other departures from what we might expect in the geometry on surfaces of non-zero curvature.

Circumference

We have already stated that the circumference of a circle on a positively curved surface will be less than 2r. For a plane (zero curvature) it will equal 2r. On a saddle, because of the ups and downs you traverse as you try to construct a circle, there will be an excess of circumference: it will exceed 2r.

Angular size

Angular size refers to the angle subtended by an object at some distance. For example, the angular size of your fist is approximately 10° when held at arm’s length. The angular size of the Sun is about ½° as seen from the Earth. (You may remember this from your Earth Science class so long ago.) The formula used to calculate such relationships in plane geometry is s = r. If I know the size and distance to an object I can calculate the angular size very quickly. Given that lines that are close to parallel converge in positively curved space, and diverge in negatively curved space, what will happen to this relationship?

s = size

r = radius qradius



In Euclidean (flat) space, s = r·.

s = size

r = radius qradius



In positively curved space, s < r·.

s = size

r = radius qradius



In negatively curved space, s > r·.

Sum of angles in a triangle

Suppose you begin a trip on the equator. You head east and travel ¼ of the way around the world. Next you go north, and don’t stop until you reach the north pole. Now return to your starting point by going straight south. You have now created a triangle. What is the sum of angles for this triangle? If you create a triangle on a saddle, what will be true about the sum of the angles there? Do you suppose there are any limits to the angle sums for either a spherical triangle or one on a saddle?

Parallel transport and geodetic rotation

Now imagine going on the triangle trip on a sphere, and you’ve got a pointer that you don’t want to rotate. (This process, of moving an object without rotation, is called parallel transport.) You point it to the east to begin. When you turn to go north, you keep the pointer pointed to the east. It stays in that direction all the way to the north pole – remember, you don’t want to rotate it. When you get to the north pole, you’ll turn 90° to go south, but you won’t rotate the pointer. As you travel south, which way will the pointer be pointing? Answer: to the north. You have completed a round trip, never

rotating the pointer relative to the surface, and yet it isn’t pointing in the original direction! This rotation is due to the curvature of the surface, and is called geodetic rotation.

**The Shape of Space**

In addition to the geometry of space, there is also the topology of space. Geometry (literally ‘measuring the Earth’) is all about measurements; topology is about connections. Although General Relativity tells us how matter affects the local nature of spacetime, it does not tell us how space is connected globally.

Normally we assume that if you headed off in a straight line you would never return to your starting point. But a sphere is an interesting surface: traveling in a straight line, you have to return to your starting point. The sphere has no boundaries, and yet it is finite. As far as connections are concerned any two points can be connected by two paths that are both straight. These are the two arcs, major and minor, of the great circle that includes the points. (You may also remember someone named Columbus who claimed he could reach the East by traveling to the west.)

A cylinder and a plane are both surfaces of zero curvature, but the space is connected differently. On a plane, there is but one straight line that will connect a pair of points. On a cylinder, there can be many straight lines. And topologists tell me that you can even have negatively curved space that circles back on itself! So although Einstein tells us much about the local properties of spacetime, we do not know how spacetime may be connected on the largest scales. For the time being, that question is left to the observers to discern.

**O’Malley and Ophelia**

Let’s see what we can say now about the worlds on which O’Malley and Ophelia found themselves. O’Malley experienced a world on which a straight line came back upon itself, and the sum of the angles of his triangle exceeded 180°. What type of world fits this? And Ophelia found that in her triangle the sum of the angles was less than 180°. The first problem in the Exercises that follow ask you to put this information together to draw the routes followed by our protagonists.

**Exercises**

2.1 What type of world did O’Malley find himself in? How about Ophelia? Can you draw a map of their travels?

2.2 Why are most lines of latitude not straight lines? Can you find a clear example of a line of latitude which is not the shortest distance between two points?

2.3 For two lines on a sphere, how many points of intersection are possible?

2.4 On a plane, if you have a line which crosses one line of a pair of non-intersecting lines, it must cross the other as well. (We call that line a transversal.) Is this still true on curved surfaces? (The answer may differ for spheres and saddles.)

2.5 What happens to the curvature of a balloon as you inflate it?

2.6 Why is the curvature of a cylinder zero?

2.7 How many straight lines connect two points on a cylinder?

2.8 Where is the curvature of the surface of a torus (doughnut, inner tube) positive, and where is it negative?

**References**

Trudeau, Richard J. The Non-Euclidean Revolution, Birkhäuser, 1987.

The MacTutor History of Mathematics archive, School of Mathematics and Statistics, University of St. Andrews, Scotland, <http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Non-Euclidian_geometry.html>

Chapter 3. Gravity and Projectile Motion

Newton’s theory of gravity involves action at a distance: two objects pulling on each other with the force inversely proportional to the square of their separation. Time does not show up in his equation; the action is instantaneous. Once physicists understood the limitations placed on forces by the finite speed of light, instantaneous actions at a distance were ruled out. A new theory was needed.

**If there is no gravity, how do projectiles move?**

We have said that an observer in freefall will simply see projectiles follow a linear trajectory. To this person, inertia is all there is to projectile motion. What about those of us standing on the ground? It sure doesn’t look like a straight path – it looks like a parabola. Einstein said that the path is straight if you look at it the right way: in curved spacetime. You shouldn’t, in relativity, consider just the spatial components of the trajectory, you should consider time as well.

Here’s a graph of what would be observed if you toss a ball straight up in the absence of gravity:

Time (s)

0

Y (m)

1

2

3

4

1

2

3

4

5

0

Notice that the path goes off the graph rather quickly. Let’s make better use of the graph by identifying each vertical mark with more than one time. When the projectile moves off the top of the graph, just transpose that spot down to the bottom of the graph to follow the motion for the next 4 seconds. (You need to be careful to get the y coordinate right *and* maintain the same slope when you do this.) You can repeat this over and over until you run out of room on the right. Then we’ll stop.

Time (s)

0, 4, …

Y (m)

1, 5, …

2, 6, …

3, 7, …

4, 8, …

1

2

3

4

5

0

This graph also applies to what an observer in freefall would observe with gravity.

But what about us? According to Einstein, we can look at this in another way. We will observe a curved coordinate system due to the presence of the Earth: spacetime has been altered. We will begin in the same way, with the same projectile at the same height and moving with the same velocity. Again, the projectile runs off the top of the graph, so we transpose to the bottom. But notice now that when you transpose, the slope (rise over run) at the top end of the first segment is no longer the same as when the projectile was launched at the bottom. That’s because of the curved grid. Let’s continue the path, following the same rules that we used above for this task:

4, 8, …

3, 7, …

2, 6, …

Time (s)

1, 5, …

0, 4, …

0

Y (m)

2

1

3

4

5

We now see that the projectile reaches a maximum value of the y-coordinate at t = 8 seconds, *and then its y-coordinate begins to decrease*. The ball will return to Earth. The motion shows an increasing y-coordinate, the rate of increase slows, and then the y-coordinate decreases more and more quickly. It is just what we calculate from Newton’s Laws, but now we have a new way of looking at what is happening.

Old way: The motion is curved in rectilinear spacetime.

New way: The motion is linear in curved spacetime – “geodesics” are followed.

In this way of thinking about things, there is not even any difference between the motion of balls and the motion of light. They merely have different slopes.

**John Wheeler: “Matter tells spacetime how to curve; the curvature of spacetime tells matter how to move.”**

Here’s a demonstration you can do at home if you’d like: Stretch a sheet of plastic wrap (such as Saran®) tightly over the rim of a wide bowl. Place two balls on the surface, one heavy (star) and one light (planet). Watch the planet roll into the star. Next, give the planet some initial velocity, and watch it orbit the star a couple of times.

Why did this happen? The planet was responding to the curvature of the surface, not to a force exerted by the star. It doesn’t matter to the planet what causes the curvature; the local curvature is all the planet knows about, and that’s what makes it move as it does.

Now try two medium weight balls on the surface, you will see them each accelerate toward the other.

The pocket each ball forms in the plastic wrap is tilted by the presence of the other ball. This tilt means the elastic surface pushes more on the outer side of each ball than on the inner side, so the net force is toward the other ball. Each ball contributes to the total deformation of the surface; each ball responds to the local deformation. This is in contrast with Newton’s explanation of why two objects attract each other: They each reach out across empty space to exert an instantaneous force on the other, a force which can be characterized but not explained. It is in this regard that we now can say that gravity is a fictitious force: there is no reaching out across empty space by one object to pull on another. Einstein has taken us a step deeper in an explanation for why gravity does what it does: Mass curves spacetime, and mass moves in response to the local curvature of spacetime. We still don’t know why matter distorts spacetime. Perhaps a quantum theory of gravity will explain this to us. There are lots of theoretical physicists working on this question. (Can you say “Nobel Prize?”)

**Space is stretched**

Go back to the demonstration with one large mass in the center. It exhibits the stretching of space, not spacetime. It shows how a two-dimensional plane of our three-dimensional space gets stretched by the presence of a mass. The plane represented is a plane containing the center of the mass causing the distortion. The stretching is greatest close to the ball and diminishes with distance. (In the graphs above, the stretching is also shown as greatest on the side of the graph closest to the Earth.)

Before we go further, what is the meaning of the radius of a circle on a curved surface like this? It’s hard to measure directly because of the curvature and the presence of the ball sitting in the middle. We make do with what is called the calculated radius: draw the circle, measure the circumference, and divide by 2. (This comes from the usual formula for the circumference of a circle.) Another way to find this calculated radius is to place a round, flat object on the curved surface, draw the circle, and then directly measure the radius of the round object. This can be done in a visual way by placing a CD or a roll of masking tape into a Vortex coin-rolling toy.

r1

r2

What are the consequences of this stretching? For one, there is more surface area to this plane than there was without the mass present. Imagine drawing two circles on the elastic sheet. The region between them is called an annulus. The formula for the area of an annulus is (r12-r22).

r1

r2

If you had a bunch of 1 cm tiles, this formula would give you an idea of how many tiles would be needed to cover a flat annulus. But with the stretching, there is truly more area between the circles than Euclid’s formula indicates. If you were to attempt to tile an annulus around our Sun, you would find that it would take more tiles than the Euclidean formula predicts. The annulus would be flat – the plane we’re looking at doesn’t really dip down and back up. But there’s more area to it than we would have expected prior to Einstein. So distances, areas, and volumes are all stretched, increased, by the presence of a mass.

How far apart are the two circles in the image above? That answer depends on how you go about finding it. One way would be to simply take the difference of the two calculated radii: r1 - r2. But another way would be to get out a ruler and directly measure the distance between the circles on the curved surface itself:

r1

r2

rmeasured

rcalculated

From the image above, you can see that the measured distance between the circles is more than the calculated distance. The mathematical formula for the relationship between these two differences will show up in the section on black holes. For now, remember that there is more space than you would have thought, and if you measure it directly you will find more space.

In particular, if you put two points on roughly opposite sides of a central mass on your elastic sheet, and then put a ribbon on the surface to show a straight line (on the surface, as in section 2 on curved surface geometry) you will find that the length of the ribbon will be longer than you’d get from the distance formula. The distance formula gives the result you would get with a meter stick measuring straight across from one point to the other, not measuring along the surface. (The ribbon, when viewed from above, will also be seen to bend a bit. This is another prediction of General Relativity and will be discussed in detail in the next section. The extra length due to this bending is small compared to the extra length due to the ribbon having to go down and back up.)

This leads to a measurable effect: A radio signal passing near the Sun will have to travel further than if space were not stretched. This means that any signal sent like this will take longer to reach its destination than Newton (or Euclid) would have expected. In going from Earth to Mars, the extra time is about 200 microseconds. Not a lot, but still measurable; the gravity at the surface of the Sun is considered weak by the standards of General Relativity. This effect has been measured and agrees with the theory. (The extra time, if due only to the bending mentioned above, would be only about 10 *nano*seconds.)

Does a beam of light really go down and back up physically as it goes by a mass? No. Every analogy is designed to show some aspect of the truth. The truth in this analogy is that there is more space close to a mass than you would expect from Euclidean formulas. Mathematically, or physically, it is not necessary to have this plane embedded in 3D space in order to achieve this stretching. It is possible that our space is embedded in a space of higher dimension, but it is not necessary. In any event, a beam of light going past the Sun, for instance, would follow a path that is fully contained by a flat plane. But it would appear to us to go slower when close to the Sun because it has to cover more distance than is apparent to us on the Earth. If you were close to the Sun watching the beam as it went by, you would measure its speed to be the same value you always observe; relative to local space, light always travels at c.

**Exercises**

3.1 Try drawing a trajectory for yourself on two graphs with different amounts of curvature. For each graph find the maximum height of the projectile. Which graph shows stronger gravity? Which has more curvature (is more distorted from a rectangle)? (The initial speed of the projectile is related to the angle the arrow makes with the axes. In each of the graphs below, the arrow makes the same initial angle.)

4, 8, …

3, 7, …

2, 6, …

Time (s)

1, 5, …

0, 4, …

0

Y (m)

2

1

3

4

5

A

4, 8, …

3, 7, …

2, 6, …

Time (s)

1, 5, …

0, 4, …

0

Y (m)

2

1

3

4

5

B

3.2 Qualitatively, how does the maximum height of the projectile depend on the amount of curvature? Explain. What would the maximum height of a projectile be if the curvature were infinite?

3.3 Is it possible to “jump off the surface” of the deformed spandex and take a shortcut getting from one point to another?

3.4 Two explorers in spaceships are orbiting a black hole. The first, Ophelia, calculated her radius to be 100,000 meters, and the second, O’Malley, finds that his radius is 150,000 meters. At a time when they are on the same side of the black hole, as shown, O’Malley puts out a tape measure to directly measure their separation. Does he find it to be 50,000 meters, more than 50,000 meters, or less than 50,000 meters? Explain.

O

O’

3.5 If you were in a free-falling elevator and you used a slingshot to shoot a pea across the elevator, would it follow a straight path (as viewed by you)? What if you shot a pulse of light?

3.6 This problem points out another problem with Newton’s view of gravity.  
  
Suppose you are in space, and you’ve got two asteroids pulling on each other. By knowing the distance between their centers and their masses, you can calculate the force of each pulling on the other. Imagine that you have rocket engines pushing outward on each asteroid to keep them at the same distance. The rockets have force measuring devices so they can show just how much force they are providing. Calculating the value of the force from Newton’s Law of Universal Gravitation gives a value in agreement with the rocket measurements. No surprise there.  
  
Next, suppose another observer shoots by at close to the speed of light.   
  
a) What will this observer say about the distance between the two asteroids   
 compared with the first astronaut?   
  
b) What will this observer say about the two masses compared with the first   
 astronaut?   
  
c) When this observer calculates the force between the asteroids (using Newton’s   
 Gravity) will the answer agree or disagree with the rocket measurements?

Note: The curved spacetime graph idea came from the book Turning the World Inside Out.

Chapter 4. The Three Classical Tests of General Relativity

There were three tests that Einstein proposed for his theory. One result was already known and General Relativity matched it well. (Does that make it a ‘postdiction’ rather than a prediction?) The other two would take more time to verify.

**I. Perihelion of Mercury**

In the 1800s astronomers were collecting more and more precise data regarding the motions of the planets. This allowed for a more precise test of Newton’s Laws of Motion as they applied to the orbits of planets. One of the basic ideas (which originated with Johannes Kepler, his 1st law of planetary motion, and was derived by Newton with his laws of motion and gravity and calculus) was that the orbits are closed ellipses.

The planet Mercury seemed to have other ideas, though. The major axis of its elliptical orbit was found to rotate (this is called precession) at a rate of about 5600 seconds of arc (aka arcseconds) per century. That’s about a degree and a half. Per *century*. (The drawing below exaggerates the effect to show more clearly what precession is.) It takes some amazing measurements to tease out an effect this small.

Sun

Mercury

This is not to be confused with precession of the equinoxes, of which you may have heard. In that effect, the axis of rotation of the planet Earth slowly turns. But that effect does come into play here indirectly, as discussed below.

The question arose, does this precession indicate a problem with Newton’s Laws? The analysis from Newton imagined only the Sun and one planet. Perhaps the real solar system harbors some less than ideal characteristics. By treating the solar system more exactly, it was thought, an explanation would be found for the precession.

By accounting for the variation in the direction of the Earth’s axis (this is the precession of the equinoxes), it was found that a little over 5000 arcseconds per century of Mercury’s precession could be explained away as an artifact of making our observations from the Earth. But still, 574 arcseconds of real precession per century were left unexplained. Since there are other planets in the solar system, it was thought they might perturb the motion of Mercury. The astronomers began calculating these effects. They found that Venus (Mercury’s nearest neighbor) contributed 277 arcseconds per century to the precession. Jupiter, while much further away, has a large enough mass to be the second largest contributor: 153 arcseconds per century. The Earth shows up as well, adding 90 arcseconds per century. Mars and the other planets together provide just 10 arcseconds per century. This adds up to 531 arcseconds per century. Pretty good, but it wasn’t quite enough. Astronomers were left with an unaccounted 43 arcseconds per century. That’s less than 1% of the original number, but it was still clearly beyond the measurement error of about 1 arcsecond per century. And they had run out of ideas.

Einstein, in trying to calculate (with General Relativity) the true orbit for a planet with Mercury’s characteristics orbiting a star, came up with 43 arcseconds per century on his first try! General Relativity had explained the discrepancy between Newton’s theory and the observations. He said, “For a few days, I was beside myself with joyous excitement.” Well, he said it in German, but that’s the gist. His theory had no adjustable parameters in it. Either the number would match or it would not; no one could claim that he had fudged somewhere in order to get the right value. This was a tremendous result, and it gave Einstein confidence that there was something really right about General Relativity. (The effect comes about because in General Relativity the strength of gravity is stronger when close to a mass than Newton’s inverse-square law predicts. That greater force swings Mercury around a bit further than would otherwise happen.)

**Experimental verification**

Mercury’s orbit provided the first evidence for General Relativity, but the precession of the other planets, although smaller, also agreed with Einstein’s predictions. But these results are all in a weak gravitational field.

In the year 1974 a new example was found which exhibited this effect: a binary pulsar, two neutron stars orbiting each other. The orbital period was only a little under 8 hours, and the precession matched predictions. Several other binary pulsars have been found since, and they all agree with the predictions of General Relativity. Most recently, a pair of white dwarf stars were found orbiting each other with a period of only 5.36 minutes! This system hasn’t been analyzed sufficiently to test General Relativity, but it will be soon.

**II. Bending of Light**

**Based on the equivalence principle**

Imagine you are out in deep space, far from any gravitating bodies. Fortunately you remembered to bring your spaceship with you. When you shoot a pulse of laser light across the ship (perpendicular to the axis of the ship), you observe that it travels in a perfectly straight line. What would happen, though, if you ignited your engine as the laser shot its pulse of light? The light would know nothing of the motion of the spaceship, so it would continue in a straight line. But the spaceship would move forward. In this way, you would see the pulse of light follow a curved path relative to the spaceship. This would simply be the effect of an acceleration of the spaceship.

Einstein’s first postulate says that the effects of gravity are equivalent to the effects of an acceleration. If we accept the first postulate, and an acceleration makes the light bend relative to the room, then so will gravity!

How much would a laser beam fall in traversing a classroom? In a simple analysis, we expect that the light would fall the distance the floor would rise in the case of an accelerating rocket in space. In a time t, that would be ½ g t2, where t is the time it takes the light to move across the room. If the room is 6 m in width, t = 6 m/c, or 2 x 10-8 seconds. Therefore the floor would rise ½ g t2 = ½ g (2 x 10-8 s)2 = 1.96 x 10-15 meters. This is approximately the size of a proton. How would you know if the beam deflected downward in the classroom by this amount? There’s no way to measure this tiny distance. But if you had stronger gravity acting over a larger distance you might have a chance.

**Based on curved surface geometry**

As a demonstration, take a cloth ribbon (they tend to be flexible and not as stiff as plastic) about 1” wide, and lay it on the elastic sheet without a heavy ball in the middle; show it is straight. Then place the heavy ball in the middle and see how the ribbon is no longer flush against the surface. Push the ribbon so it conforms to the sheet as it passes near the heavy weight, and the ribbon will clearly be going a different direction after passing the “star.” A laser level mounted above will show what a straight line (to us) would be, and the ribbon clearly does not follow the laser’s path.

The stretching of space by a mass will cause a change in direction of a light beam. The full bending expected by relativity isn’t apparent, though, since half of the bending comes from the time distortion (as in the curved spacetime demonstration in an earlier section).

**Experimental verification**

Einstein saw that it would take much stronger gravity than at the surface of the Earth to bend light enough to be measurable. The strongest gravity in the solar system is at the surface of the Sun, so he calculated just how much a beam of light would be deflected if it passed near the Sun. He derived the formula

Sun

r





which gives a value of 1.75“/ r, where in this simplified version r is measured in solar radii. (The symbol “ stands for arcseconds.) This would be just measurable at the time, but with one problem. You’d need to look for a displacement of a star in the sky right next to the Sun. It’s a little hard to see stars in the daytime. But during a solar eclipse the Sun is hidden, and stars do come out. So plans were made to observe the next available solar eclipse, scheduled for August 21, 1914. But World War I intervened and the observations were canceled.

The next solar eclipse to be useful was in May of 1919, and this one would be right in the middle of the Hyades cluster (whose stars make up much of the constellation Taurus). There would be lots of bright stars surrounding the Sun, making this an excellent opportunity. When the results were announced, Einstein’s predictions were found to be in agreement within the experimental uncertainty. It was this prediction that really put Einstein in the public eye, making him a personality known beyond the physics community.

Modern measurements give an accuracy of 100 microarcseconds, and Einstein’s prediction of 1.75 arcseconds still holds up. The deflection by Jupiter is calculated to be only about 300 microarcseconds, and agreement was found (although you can see that the uncertainty in measurement is a big fraction of the measurement itself).

**Gravitational Lensing**

The bending of light was discovered on a quite large scale as well. Back in the 1960s astronomers discovered bright point-like objects with spectra that made no sense at all. They had lines, but the lines didn’t seem to correspond with any elements. Eventually someone saw that the lines did match up, but not with visible spectra – the lines were what should have been ultraviolet lines of hydrogen, but redshifted so much that they showed up in the visible part of the spectrum. These objects would have to be extremely far away to have such a large redshift, and that meant that for us to be able to see them they would have to be putting out much more energy than any known star. They were named “quasi-stellar objects,” or quasars for short.

Soon many examples of quasars were found, each with its own spectrum. Well, just about. There were a couple that had the same spectrum, and they also happened to lie very near each other in the sky. With more careful measurements it was found that there was a galaxy lying between us and the quasar that bent the light from the quasar around both sides of the galaxy, yielding two separate images of the same quasar.

Galaxy

Quasar

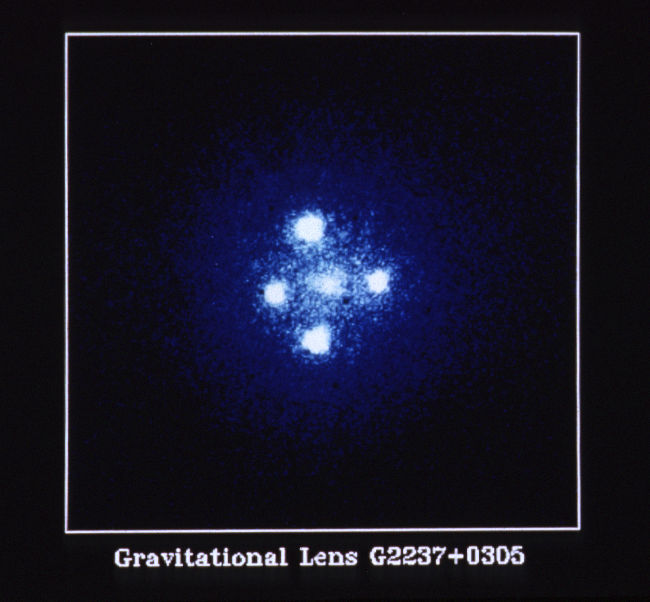
Earth

Image 2

Image 1

If the three objects (quasar, galaxy, and Earth) are lined up well enough, you can even get an “Einstein ring” where the quasar image looks like a ring around the galaxy! More commonly, gravitational lensing distorts the appearance of the distant object.

In the following image, taken by the faint object camera on the Hubble Telescope, a quasar 8 billion light years away shows up four times, its light bent by the gravity of a galaxy (in the middle of the image) only 400 million light years distant. It is sometimes called the “Einstein cross.”



The next image shows all sorts of streaky things. There is a galaxy cluster that is relatively compact, so its gravity is strong and bends light quite a bit. Those streaky things are galaxies that are much more distant that the cluster. If it weren’t for the gravitational lens provided by the cluster, those distant galaxies wouldn’t be visible to us at all. Their images may be highly distorted, but at least we can see them.

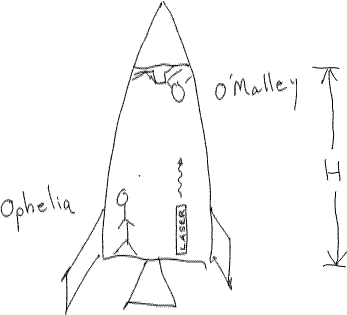


The bending of light by the curvature of spacetime is well enough accepted that it is used to map out the presence of matter, even when that matter doesn’t emit enough light to be seen. In particular, this technique is used to determine the amount and location of dark matter in our universe.

**III. Gravitational Redshift**

The final classical test of General Relativity proposed by Einstein is that light traveling upward in a gravitational field would suffer a shift to a lower frequency (termed a redshift). How does this work? We’ll use the equivalence principle again.

Imagine Ophelia and O’Malley in a spaceship in deep space, well away from any significant gravitating bodies. Ophelia is on the bottom, as shown, and O’Malley is at the top, a distance H away from Ophelia. (Without gravity bottom and top are pretty meaningless, but I hope the illustration makes it clear.) Ophelia has a laser, and she shoots a pulse of laser light up toward O’Malley. The engine is off, and when the pulse reaches O’Malley he measures the frequency. Not surprisingly he finds the same result as Ophelia does when she measures the frequency.



If the ship is accelerating upward, though, the answer will be different. In this case, O’Malley will need to don his Spidey suit to hold himself in place on what is now the ceiling. The pulse takes a little time to reach O’Malley, and in that time he will have accelerated just a bit. He will be going faster when he receives the light than the laser was going when it fired the pulse. He and the laser are still together in the ship; they are going the same speed. But since both are speeding up, and the pulse is produced by the laser before it is measured by O’Malley, the two events happen at different times. Since O’Malley is going faster than the laser was when it emitted the light, it is as though O’Malley is going faster than the laser. It’s the same effect as if he were running away from a stationary laser.

If you are moving faster than a source of waves in a direction away from the source, those waves will be Doppler shifted to a lower frequency. So the light O’Malley sees will be Doppler shifted: he will measure a lower frequency than Ophelia does. Qualitatively, that’s the prediction.

The final step is to use the equivalence principle: If an acceleration causes a redshift, then so will gravity. There will be a gravitational redshift.

Can we find a formula ourselves? An approximate result is not too hard to find. The frequency will be found with the Doppler formula for light1:



In this formula, u is the relative speed of the source and observer, with positive values indicating that the source and observer are moving away from each other. The speed of light, c, and the frequency of the source, fsource, are parameters in this problem. We need to find the value of u to plug in.

How much faster will O’Malley be going by the time the light reaches him? We will assume constant acceleration, so

,

where “t” is the time it takes the light to travel from Ophelia to O’Malley. As long as the spaceship is going at modest speeds (much less than the speed of light) we can approximate the time with the simple expression



Therefore .

We can now plug this into the formula for the Doppler shift:



If we divide both the numerator and the denominator inside the square root by c and substitute , we get

.

Given an acceleration of 9.8 m/s2 and a value of H attainable in a building, we can see that  is very small. That allows us to simplify using a Maclaurin series. We’ll just keep the first two terms:



Finally, we get our result:

.

The shift in the frequency is given by

.

For the gravitational version, we simply use ‘g’ for the acceleration.

Consider the term multiplying the source frequency on the right-hand side. For an experiment where a = g and the height through which the light travels is on the order of 20 m (about a 7 story building), we get a number on the order of 10-15. The frequency shift is a tiny fraction of the original frequency. It took until 1959 (40 years after the last test of General Relativity) for this to be measured. It was done by Pound and Rebka, who used the old Harvard Tower (22.6 m) for the test. The predicted fractional shift (the multiplier of the source frequency in the above equation) was 2.46 x 10-15, and the measured result was 2.57 (+- 0.26) x 10-15. It was an amazing result.

1The Doppler formula for light is different than the Doppler formula for sound waves. That is because sound waves travel at a fixed speed relative to the medium in which they are traveling. For light waves, relativistic time dilation, small though it may be at low speeds, compounds the frequency change due to the motion of O’Malley.

**Exercises**

##### Perihelion Problems

4.1 Why would Mercury show the largest perihelion shift (of all the planets) due to General Relativity?

Bending of Light

4.2 Suppose you have two stars in the night sky. Also suppose that six months later the Sun is directly between those two stars (because the Earth is now on the other side of the Sun). Will the two stars appear closer together or further apart now that the Sun lies between them in the sky?

4.3 Suppose you had three spacecraft surrounding a black hole. The astronaut in each ship points a laser beam so that it hits the next ship. The laser beams form a triangle of sorts. How will the sum of the angles of the triangle compare to 180 degrees?

4.4 What is the maximum bending of light as it passes by the Earth?

4.5 Suppose you wanted to use the bending of light by the Sun as a telescope lens. If a source of light was many light years away, where would the light rays passing on either side of the Sun converge?

4.6 If a massive object, such as a large planet, passes in front of a distant star, the brightness of the star can be increased temporarily. Explain how this can happen. (The effect is called microlensing.)

Gravitational Redshift

4.7 What is the fractional change of frequency (f/f) of light leaving the surface of the Sun and heading out to infinity? What about light leaving the surface of the Earth?

4.8 Suppose that O’Malley, at the top of the spaceship, has the laser and points it “down” toward Ophelia. What will she have to say about the frequency of the laser? Why?

4.9 Cosmological redshifts (due to the expansion of the universe) have been observed that produce fobserved values that are as small as ¼ of fsource. How does the gravitational redshift from the surface of a star compare to this?

4.10 If light of a frequency of 5 x 1015 Hz is emitted from a radius of r = 2GM/c2, find the frequency of that light at infinity. What do you suppose this answer means?

Chapter 5. Gravitational Time Dilation

***The existence of the gravitational redshift implies gravitational time dilation.***

Ophelia and O’Malley have their laboratories on different floors of a building. Ophelia is on the ground floor and O’Malley is a few floors up. Ophelia, being the clever physicist she is, has built a new clock based on light waves.

**Wave basics, not limited to relativity**

The Greek lower case lambda, , is used to represent the wavelength of a wave. It is the distance between crests. We use the letter v to represent the velocity of a wave.



x axis

Imagine this wave moving along in the positive x direction, like waves on a lake moving past you in your canoe. After just a little bit of time, one crest will have moved forward to the spot that the crest in front of it used to be. (The dashed wave indicates the original position of the wave.)

t = 0

t = T

.

.

.

v

If you had closed your eyes in between these two times, you wouldn’t know that the wave had moved at all. (Except that waves don’t just sit still. People can’t water ski by skiing down stationary hills of water.) The time it takes for one crest to move to the position of the next crest is called the period of the wave; we use the letter T for the period. (T for time.) If your canoe were sitting still, you would see one crest go by every T seconds. So the wave has moved a distance of  in a time T moving with a velocity of v. These variables are related to each other: distance equals rate times time, or  = v · T.

One more variable we use with waves is f, for frequency. The frequency is the reciprocal of the period, so since the period is the number of seconds per crest, the frequency is the number of crests per second. The units are inverse seconds, s-1, also known as Hertz (Hz). This gives us another way of expressing the earlier equation with frequency instead of period: f ·  = v.

**Ophelia’s clock**

The clock involves a laser that oscillates (as all lasers do) at a very precise frequency. She put a piece of glass into the beam to deflect a small fraction of the beam to a very fast counter. It counts the crests of the light waves from the laser. And since the laser has a well known frequency, it is possible to count just the right number of crests to equal one second. For this laser, it is one quadrillion (1015) crests. The counter is set to make a “tick” sound every time it counts 1015 crests, so it goes “tick, tick, tick, …” just like a mechanical clock. Here is a schematic of the clock that Ophelia built:

beamsplitter (glass)

laser

crest counter

speaker

Since her laser only uses part of the beam, the rest will continue on its way. She has set up her clock in a stairwell so that the remainder will travel straight up to O’Malley to give him access to the beam as well. And being the nice person that she is, Ophelia has constructed two counters, and has given the second to O’Malley. So he’s got a clock as well, and the rest of the beam isn’t wasted. (Much of the expense of this clock was in building a laser that was steady enough in its frequency to make a really good clock, so you don’t want to waste the extra beam.)

O’Malley set up his clock and found that it was slow in comparison to his own clocks. Not by much, but it wasn’t right. He sent the unit back down to Ophelia, who put it next to her own clock, and they kept perfect time. She sent it back.

Again, O’Malley found the clock to be slow. We know why, based on the previous section: the light from the laser traveled upward through the gravitational field and as a result was shifted to a lower frequency as it traveled up to O’Malley’s floor. The frequency as measured by O’Malley is just a little bit, let’s say just 1 Hz, lower than the frequency on the ground floor. So O’Malley has to adjust his counter so that it will tick every 1015 - 1 crests. Okay, but so what?

Here’s what. If they agree to let the clock run through 3600 quadrillion cycles and then go out for lunch, Ophelia will say that exactly one hour has gone by. O’Malley will agree that 3600 quadrillion cycles have past, but by his clock that is just slightly more than one hour. While the clock went through 3600 quadrillion cycles, Ophelia and O’Malley have experienced different amounts of time passing.

This is not the same as in Special Relativity where each observer sees the other’s clock running slowly. Here, if we put the laser on O’Malley’s floor and shoot the beam downward, Ophelia will see a blueshifted beam that oscillates too fast, according to her measurements. She will say that O’Malley’s clock is running too fast.

Conclusion: Clocks that are deeper into a gravitational field will run more slowly.

**The exact formula**

The period of a clock could be related to the period of a wave, so since the period and frequency are reciprocals, the formulas are nearly identical. If we also add in the effect of velocity, we find:



This formula gives the period of a clock near a mass M (and moving at a speed u) *as it would be measured by another clock at infinity (and at rest)*. With the denominator less than one, the period of the close clock will be measured to be greater than the period of the same clock located at infinity. (The variable rcalculated is the calculated radius of the location of the clock near the mass M.) If the period is larger, that means it is running slowly. And slow clocks measure shorter time intervals than they should. The formula for measured time intervals is

,

just the reciprocal of the relationship of the periods.

**Experimental verification**

Time dilation has been verified experimentally. The first direct test came in 1971: Two atomic clocks were used, one on an airplane and one on the ground. In this experiment both the velocity and gravity induced effects are present. The measured difference in the duration of the trip was small, on the order of 100 nanoseconds, but it agreed with the predictions of General Relativity. (It was even done for both east-going and west-going flights around the world. Because of the rotation of the Earth planes going in these directions have different velocities as measured by an inertial observer. Both flights agreed with Einstein’s theory.)

There is a modern application of this effect where even these small differences can become critical. GPS (Global Positioning System) works by finding how far away you are from several satellites orbiting the Earth. How is your distance found? Through the use of our favorite equation, distance equals rate times time. Radio signals are electromagnetic waves that travel at the speed of light, so if you know how long it takes a signal from the satellite to reach you, just multiply by the speed of light and you’ll have your distance from that satellite.

Once you know how far you are from one satellite, you know you are somewhere on the surface of a sphere centered on the satellite with a radius equal to your distance. Now you contact another satellite and find another distance. Now you know you are also on the surface of a new sphere. If you’re on both spheres, you must be located somewhere on the intersection of the two spheres. (Generally the intersection of two spheres is a circle.) If you can contact one more satellite, your position will be narrowed down to the intersection of a sphere and a circle. That’s two points. One usually is way off the surface of the Earth, and the other is near it, so you can guess which one you’re at. (Usually the GPS receiver contacts a 4th satellite as well to reduce the uncertainty in the result.)

Light travels very quickly, so getting the timing right is critical. To know how long it takes a “tick” from a satellite to reach you, you have to know when the satellite sent the tick. For this to work the satellite has to have a very accurate clock. If the speed of the clock is off by only a nanosecond (10-9 second) per second that goes by, that would throw off your distance by 3 x 108 m/s x 10-9 s = 0.3 m for every second that goes by. After 10 seconds, your distances would be off by 3 meters, and after an hour your position would be off by a kilometer. By the time one week has passed, peoples’ positions would be off by close to 200 kilometers. This would not be a very good system; you could only trust it for the first few days after launch of the satellites. The effects of General Relativity were calculated ahead of time, and the clock speed for the satellites was set in advance to correct for this.

**Exercises**

5.1 If you live 100 years on the surface of the Earth, how long would an observer at infinity claim you had lived? (You should ignore the existence of the Sun in this problem, and just calculate the effect due to the Earth’s gravity.) Hint: A Maclaurin series may help you here.

5.2 To find the relationship between the periods of two clocks, both at finite locations, you can do the following: Write out the formulas comparing the period of each clock to a clock at infinity, and then use these two equations to eliminate the period of the clock at infinity. What is the resulting formula?

5.3 GPS (Global Positioning System) satellites travel in LEO (Low Earth Orbit) at an altitude of about 400 km. Compared to a clock on the surface of the Earth, by what factor will the clock be off? Will it be slow or fast relative to the Earth clock? (The effect of its speed will slow it; less gravity than on the surface of the Earth will speed it up.) Note: This effect was calculated in advance because it is big enough to keep the GPS system from working properly if not accounted for.

5.4 If you wanted to live a bit longer, should you live at the ground floor of a building or in the penthouse? For this problem, consider only the effect of gravity and ignore the velocity of each apartment.

5.5 If you lived at the top of a 100 meter tall building and another person lived on the ground floor, after 100 years how much would your clocks differ by? (Ignore the velocity due to the rotation of the Earth in this problem.)

5.6 If you wanted to live a bit longer, should you live at the ground floor of a building or in the penthouse? This time, remember that the Earth is rotating. The height of the floor you’re on affects your speed, since you are further from the center of rotation if you’re in the penthouse. Does your answer to this question depend on the height of the building?

Chapter 6. Black Holes

**Origins**

The beginnings of the idea of a black hole came about in the late 1700s, long before Einstein developed his theories. Both John Michel and Pierre Laplace independently took the formula for the escape velocity from a spherical object and plugged in c for the escape velocity. This gave a formula for how small such an object would have to be in order for light to be unable to leave the surface.

; 

For the Sun this radius is about three kilometers, implying a density of matter higher than in a nucleus. The equations came from Newton, and they paint a picture which is incorrect (for example, light would not rise up from such a surface and then fall back), but oddly enough they give the correct value of the radius for a black hole.

The relativistic result which characterizes black holes was derived by Karl Schwarzschild, a physicist who died during World War I. (His name, oddly enough, means “black shield,” which in relation to black holes sounds like the event horizon, which will be defined later.) Schwarzschild produced a formula for the stretching of spacetime due to the presence of a spherically symmetric, non-rotating point mass.

Because you can’t stretch a tape measure from the center of a black hole out to where you might be located, it isn’t possible to directly measure your distance from the center of the black hole. As a result, you need to resort to calculating your radius from a circumference. Whatever your radius is, imagine a circle of points (including your location) in which the black hole is centered. Using the standard formula for circles, you can calculate a radius

.

This is the same method we used with the Vortex toy and the elastic sheet back in Chapter 2. The presence of the mass of the black hole stretches space (and curves time, too, but we won’t be worrying about that right now). This means that if you move toward or away from the black hole, you will be traveling along the surface of the stretched space. As discussed in section 3, you will travel farther in going from one point to another than you would have guessed based on the change in your radius. As a reminder, the two changes in radius are graphically related as follows.

r1

r2

rmeasured

rcalculated

r1 and r2 are calculated radii.

The formula relating the two, derived by Schwarzschild, is

.

Remember that the measured radius to the center of a black hole is not something we can directly measure. Notice that the denominator goes to zero as the calculated radius approaches the value found originally by Michel and Laplace. It is now known as the Schwarzschild radius.

Note: In this formula, you can switch dr (an infinitesimal change) to r (a finite change) so long as the denominator does not vary much over the range of radii under consideration. Otherwise you would need to integrate to find the change in radius. I would recommend a table of integrals for this if you need to do this.

If our Sun magically turned into a black hole today, with all of its mass packed away into a region with a radius less than a few kilometers, the orbit of the Earth would not change. Out at our distance, all that matters is the amount of mass curving spacetime, not how compact it is. All the really weird effects of black holes happen up close, within a relatively small number of Schwarzschild radii where the denominator begins to vary noticeably from one. For our Sun, that would be within 1000 km or so. The Earth is roughly 150 *million* km from the Sun. Black holes don’t really act like stellar vacuum cleaners, sucking in all that surrounds them. When you think about all the predicaments science fiction protagonists find themselves in regarding black holes, you can see that they must be very unlucky to come so close to something so small in all the vastness of space.

**Why are black holes black?**

In the Newtonian view, an object launched from the surface of a planet with less than the escape velocity will rise, slow down, and fall back to the ground. On this basis, you might imagine that a pulse of light sent upward from a black hole would do the same thing. This was what Michel and Laplace had in mind.

But in Einstein’s way of doing things, light always follows a straight line (in curved spacetime). And it must always move at c relative to any observer. Back when we used the curved graph paper to plot the trajectory of a projectile, we saw that the curvature of time was largely responsible for making a projectile fall back to the Earth’s surface. What would happen if the curvature were greater than the case we examined? The ball would reach its maximum faster. With sufficient curvature of both space and time near a mass M, any geodesic will head toward the center: The light will never move outward from the Schwarzschild radius. Any light emitted from the Schwarzschild radius (or inside it) will be unable to move outward at all, even a little bit. This is why the sphere defined by the Schwarzschild radius is called the *event horizon*. A horizon is something you cannot see beyond. Any event that happens inside the event horizon will never be seen by those of us on the outside. And if you venture inside, you’ll never come back out again to tell us what it’s like.

Weirdly enough, a beam of light shot straight outward from the exact Schwarzschild radius would seem to stand still from an outsider’s point of view. This is not a violation of the speed of light postulate from Special Relativity. Einstein’s rule about light is that any light moving past an observer will be seen to travel at c. If the light stands still, it isn’t moving past you. What would a stationary observer at the event horizon say? According to Einstein, it is not possible to have a stationary observer at the event horizon. An observer at the event horizon would be falling inward at c, and so a beam of light standing still would appear to be moving outward at c. It is as though space is a conveyor belt moving toward the center at the speed of light, so that the light can be moving outward at c relative to space and yet not get anywhere.

Black holes (named by John Wheeler, a really neat relativist) are very democratic, too. They will incorporate anything that happens to fall into them. From the outside, the only properties of a black hole we can measure are its mass, its electric charge, and its spin (angular momentum). Nothing else matters. It doesn’t even matter if the black hole were formed from matter or antimatter. As far as we can tell, black holes only have these three quantities that are measurable. It is sometimes said that black holes “have no hair,” nothing to distinguish one from another, beyond these three quantities.

**Tidal effects: the dark side**

Earlier you learned that gravity from sources such as planets will vary depending on your distance from the center of the planet. We will look at a simple case where two balls, tied together with a piece of string, are freely falling in the gravitational field of some spherical object.

M

2

1

m

m

To evaluate the motion of each object we will start with free-body diagrams:

Applying Newton’s Law of Universal Gravitation (it works pretty well even if it isn’t quite right) and Newton’s 2nd Law of Motion:

m

m

Fg1

Fg2

T

T

1

2

 and .

Since we intend for these objects to stay together, we will assume that the accelerations are equal. As a result, the two sums must be equal:

.

Solving for T, we find

.

If you have a small separation compared with the distance to the gravitational source (if

d << r), this can be approximated as follows:

.

Since  is small, we can use the Maclaurin series approximation: .

This puts the tension in the string into the following form:

.

This is the tension needed in the string to hold the two balls together in the varying gravitational field due to M. Normally this isn’t a problem since the closest they could get to the Earth is to be near its surface, which makes r equal to about 6 million meters. As a result, the tidal effect on them is just not noticed. But if you compacted the Earth to a black hole, they can get a lot closer. As r gets smaller, the tension rises quite rapidly. If they got too close to a source of gravity, this tension could be stronger than the string could withstand. In that case, the string would break and the balls would accelerate at different rates. If you know the maximum tension the string can withstand, you can solve for the closest distance to M that the string can handle:

.

(This effect shows up not just in the tides, but also in one of the most beautiful things you could see in the solar system: the rings of Saturn. Within a certain distance called the Roche limit the gravity that normally would hold a bunch of chunks of ice together as a small moon is too weak to keep them together. A moon could not form within this distance. That’s why Saturn’s rings are there – the tidal force overcomes gravity for those orbiting chunks of ice, keeping them from clumping together.)

Let’s try this out for the Sun, the biggest source of gravity around. If we plug in the mass of the Sun for M, take d to be something small, say 1 meter, m equal to 1 kg for each ball, and make Tmax = 100 N, we get r at just over 1 million meters, or 1000 km. But the Sun’s radius is close to 700,000 km, so it will not be possible to get close enough to have tidal effects pull the balls apart, even for the Sun. But let’s pretend the Sun’s mass were in the form of a black hole, of radius 3 km. If we took the two balls and released them so they could fall toward the Sun, the string would break once the balls got to a distance of 1000 km. If you tried to fly a spaceship into a black hole of this size, it would be pulled apart before it got to the event horizon.

What about bigger black holes? Since the Schwarzschild radius grows linearly with the mass and the limiting radius grows only as the cube root, these two numbers will get closer to each other as M gets bigger. At a sufficiently large value of M, the limiting radius (at which an object would get pulled apart) will be inside the event horizon – an object will not necessarily be ripped apart on its way into a black hole. (It will get ripped apart, just somewhere inside the event horizon where we can’t see it happen.)

As an example, suppose M is one million times the mass of our Sun. The minimum safe radius for the balls will grow by 100 (the cube root of 1 million) to 100,000 km, and the Schwarzschild radius of the black hole will be 3,000,000 km. The tidal force from this black hole will not tear apart the balls until they are well inside the event horizon! They can enter the black hole intact. *And so could you.*

**What would it be like to visit a black hole?**

Ophelia and O’Malley have decided to take a well-deserved vacation to the Black Hole Hotel. It is a hotel that has floors that reach down to near the event horizon of a hugely massive (about a million solar masses) black hole. The gravity may be a bit strong, but hey, it was a free trip as long as they sat through the time-share condo sales pitch.

The floors are 1 km apart, and they begin at a calculated radius of 1 kilometer above the event horizon up to a calculated radius of 501 km above with new floors every 1 km. The floors are numbered by their calculated radius. Ophelia was given a room on the 50th floor (50 km above the Schwarzschild radius) and O’Malley got a room on the 200th. They agreed that after getting settled in their rooms they would meet for dinner. It was 3 p.m., so they decided to meet in 3 hours at the restaurant on O’Malley’s floor at 6.

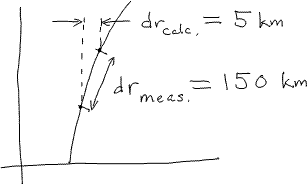
O’Malley cleaned up, took a nice nap, and got up at 5. He changed into more formal wear for dinner, and went to get a table. He arrived about 10 minutes early, got a table with a good view of the event horizon, and waited for Ophelia. He ordered a couple of diet colas and some nachos. He drank his cola and still Ophelia hadn’t arrived. He finished the nachos, and still Ophelia hadn’t arrived. He drank her cola, and she was still absent. Eventually he decided to go find her. No problem, he thought: I’m on the r = 200 km floor, and Ophelia is on the r = 50 km floor. I’ll just go down 150 km and knock on her door.

About 6:30 O’Malley left the restaurant and went in search of Ophelia. He found the stairwell, and measuring the first flight found its length to be 1 km. Peeking out the door he saw the next floor of the hotel. “This’ll be easy,” he thought. “Each flight of stairs is one floor. I’ll just go down 149 more flights and find her.”

After an hour or so, he had gone down the full 150 flights of stairs. It was now 7:30 by O’Malley’s watch. Once off the stairs, he headed for the corner room, which was Ophelia’s. He knocked on the door; an old man answered, shouting, “Have you finally brought my Ovaltine?” O’Malley apologized for the error and went to the elevator. A sign there said that he was on the 195th floor. How could this be? What did this mean? How could he be only 5 km down from his own floor after going down so many steps?

There was a phone by the elevators. (“Why didn’t I take the elevator in the first place?”) He asked to be connected to Ophelia’s room. When the phone was picked up, a strange voice (but oddly familiar) said hello. “Who is this?” O’Malley demanded. “Tthhiiss iiss Oopphheelliiaa.” The voice was pitched an octave low, and was slow. “Wwhhoo iiss thiiss? Yyoouu wwookke mmee uupp ffrroomm aa nniiccee nnaapp. Aanndd wwhhyy aarree yyoouu cchhaatttteerriinngg aatt mmee lliikkee aa cchhiippmmuunnkk?”

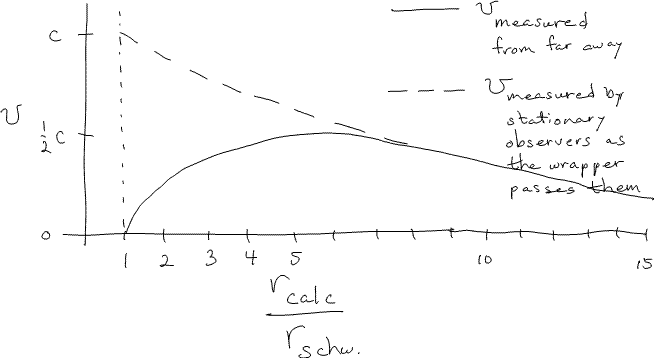
Then it hit him. Very near to the surface of a black hole, time slows down as the square root of the altitude above the event horizon. (A Maclaurin series applied to the time dilation formula will show this, for altitudes that are small compared to the Schwarzschild radius.) Since Ophelia was about one fourth of the calculated distance to the event horizon that he was, her time was passing at about half the rate of his. It was now 8:00 p.m. by his watch; 5 hours had gone by according to him. Ophelia had experienced only 2½ hours, though, so for her it was only 5:30. And he also understood how he could have gone down so far and not reached Ophelia’s floor: the floors were numbered by the calculated radius, and he had gone down a measured distance. This close to the event horizon space has been stretched a lot, so a small change in the calculated radius turns into a big change in the measured radius. He just hadn’t gone down far enough.



The worst of it was that after just 5 hours near the 200th floor, 80 days had passed back home on Earth! His library books were going to be way overdue, and he didn’t even want to think about what he had left at home in the refrigerator.

**Are there any other weird effects?**

What if you dropped the wrapper to your candy bar off the edge of the observation deck at the top floor of the hotel? It would begin to accelerate toward the black hole, and as it gets closer the acceleration will increase, as expected. If it happened to fall past the balconies of rooms on its way down, observers on different floors of the Black Hole Hotel would report increasing speed as it passes lower and lower floors, just as you would expect. These observers find the velocity by taking drmeasured/dt, with dt being measured by their own watches. Making a graph of the reported speeds vs. the calculated radius, you find the following result:



Note that the wrapper’s speed approaches the speed of light as the wrapper nears the event horizon, according to stationary observers at the various radii.

But what will *you* see? You will measure the speed as the rate of change (according to your watch) of the calculated radius. As the wrapper approaches the event horizon, space is stretched more and more. So to you (a person who sees a calculated radius) it looks like the wrapper speeds up initially when you drop it, but as it moves into more severely stretched space it slows down according to your measurements. Just as the measured distance between floors of the hotel is greater than the calculated distance, the candy wrapper travels a greater distance than you are able to directly ascertain from your location. Your disagreement with the hotel residents is also due to the fact that, according to you, the observers in the lower levels of the hotel have slow watches. With your watch ticking away seconds faster than theirs, you will observe the wrapper to move this smaller distance in a larger t. Since the time dilation becomes infinite at the horizon, the wrapper doesn’t seem to ever quite make it to the event horizon by your observations. But the gravitational redshift guarantees that the light from it will gradually be shifted right out of the visible spectrum. It will disappear from sight as it approaches the event horizon. In a sense the wrapper will never go through the event horizon according to your observations, but it will disappear into the black hole nonetheless.

What would it be like for you, if you fell into the black hole? No bumps in the road: you’d sail smoothly through the event horizon. There is nothing there to tell you where it is. But people would stop responding to your emails (the signals don’t get out). And there is a stranger thing that happens. In the everyday world, you can move back and forth in spatial dimensions, but you can’t do the same thing in time: you always move forward. Inside the event horizon, space takes on this timelike characteristic: You find that you will move toward the center of the black hole, and no firing of your engines can stop your progress toward the center. Here’s a mind bender: If you pointed your flashlight outward as you fell toward the center, when inside the event horizon, what would happen? The light would indeed move away from you at the speed of light, just as you would expect. And yet we said that light can’t move outward. What’s going on? Since you were traveling at a velocity approaching c as you neared the event horizon, what do you suppose has happened to your velocity once you’re on the inside? No one on the outside can see what you’re doing, so no one can say they have measured you moving faster than light. This is strange. Relative to space you are not moving faster than light; but space is acting like a conveyor belt, pulling you inexorably to the center of the black hole.

**How do black holes form?**

With Einstein’s theory of gravity, a number of astrophysicists began to wonder about the fate of large stars. For many stars in our galaxy, when the hydrogen that supports fusion in the core is used up the star will gradually collapse. They will still support hydrogen fusion in a shell surrounding the core and fusion to carbon in the core itself. These stars are called white dwarfs, and are very hot. Gradually, though, they will completely run out of fuel. At this point, they will have turned into lumps of cold carbon. The electron clouds of the atoms experience a quantum effect known as “degeneracy pressure” that keeps the atoms from collapsing any closer together. This is to be the fate of our Sun. At that time it will be smaller than the Earth!

But it was shown by Chandrasekhar that if a star’s collapsing mass (what is left after the star erupts a great portion of its mass into space during its red giant phase) exceeds 1.4 solar masses, the electron degeneracy pressure is insufficient to support the mass of the star against the great strength of gravity. In these stars, the electron clouds would collapse, and the electrons would combine with the protons in the nuclei to form a sea of neutrons (with the concomitant emission of a huge number of neutrinos). What you get is called a neutron star, where gravity is counterbalanced by the neutron degeneracy pressure. Some rotating neutron stars with magnetic fields give off pulses of electromagnetic radiation that shine through the universe like the beam of a lighthouse. These are called pulsars. Many have been observed. The original mass of these stars was more than eight to ten solar masses.

But if the star begins with something greater than 20 solar masses, the collapsing core will be so massive that the neutron degeneracy pressure will be insufficient to support the mass of the star against gravitational collapse. The core will continue to collapse past nuclear density, and there is nothing that will stop it from forming a black hole. (There is speculation that there may be one more type of star between the neutron star and a black hole: a quark star, in which the neutrons have dissolved into their constituent quarks, and what you’ve got is a sea of free quarks. There is a smidgeon of evidence supporting this, but it’s still pretty speculative.)

What happens to the mass that causes the black hole? According to Einstein’s equations, the matter inside a black hole will follow a path that takes it to the very center. All that mass goes down to a geometric point. Since at this point density becomes infinite (and other quantities do, too) this is called a singularity. Mathematically this means that the functions are not well defined at this point. The occurrence of a singularity in a theory is generally considered a flaw. There are several ways around this problem, though. Perhaps the slickest but least savory answer is that since the event horizon is between us and the singularity, we never can see what happens. We are saved from seeing the singularity by the event horizon. Not a very satisfying answer. Another answer is more satisfying but currently incomplete: Due to quantum effects, the wave function of the matter at the center of a black hole is spread out. (To be a point would violate the Heisenberg Uncertainty Principle.) The problem with this is that at the center of a black hole, both gravity and quantum mechanics would be important, and we need a theory of quantum gravity in order to model what it would be like. And we don’t have a theory of quantum gravity.

Einstein did not believe that black holes could form in the universe. He believed that something would have to stop the matter from reaching a singularity, but there were no forces that could withstand the force of gravity in this case, as shown by his equations. Since he didn’t think that all matter could go to a point in the center (the singularity) he didn’t believe that black holes would be allowed.

One other method is possible for the creation of a black hole, but we don’t know if it has happened in the universe. This method may have come into play for the largest black holes known – those at the centers of galaxies with masses equal to millions (even billions) of stars. The density of matter needed to form a black hole would be the mass you’ve got divided by the volume within the Schwarzschild radius. Since volume goes up as the radius cubed and mass is proportional to the Schwarzschild radius, the required density goes down as the mass rises. So what? Well, the density you would need to achieve with all that mass isn’t the nuclear matter we’ve talked about earlier. With a mass approaching a billion solar masses, the density you’d need for the formation of a black hole would be less than the density of water! The idea here is that in a huge early galaxy, perhaps the central bulge of gas simply contracted sufficiently that (without having to overcome the nuclear repulsion) a huge black hole appeared without any need for supernovas. (It is also possible that the huge black holes in the center of galaxies grew out of collisions of normal sized black holes, gradually building up to their current sizes.) We need more detailed observations of the earliest galaxies to resolve this issue.

**Can a black holes ever die?**

In thermal physics, the only objects that cannot radiate energy are objects at absolute zero. Since it was believed nothing could escape a black hole, it was thought they would be at absolute zero. But this would allow for the entropy of the universe to decrease when things fell into a black hole. No one liked this conclusion, and finally Stephen Hawking found that black holes have a finite non-zero temperature. Therefore they should radiate. The explanation required the use of Quantum Mechanics as well as General Relativity. In Quantum Mechanics, particle/antiparticle pairs can pop into existence as long as they disappear again quickly. Hawking showed that such a pair forming near the event horizon could lose the negative energy member of the pair through the event horizon, providing the energy needed for the positive energy member of the pair to escape to infinity. (An alternate explanation from Hawking is that a particle inside the black hole could “tunnel” out, following the rules of Quantum Mechanics.) This quantum effect would allow radiation to leave the black hole. The formula derived by Hawking gives the rate of loss of mass (in kg/s) using meters, kilograms, and seconds for any values needed, is



When you plug in the values for the various constants, this simplifies to



But how noticeable would this be for stellar-mass black holes? The mass of our Sun is roughly 2 x 1030 kg, and so when you square that in the denominator you get a tiny result. Since the rate of evaporation of a black hole by this process depends inversely on the mass squared, a small black hole would be radiating much faster than this. Therefore, as a black hole radiates and loses mass, it will radiate faster and faster until it finally disappears in a bang, a burst of gamma rays at the end. But given how slowly a big black hole radiates initially, it would take a long time to reach the end. The lifetime of a black hole that starts out with the mass of our Sun would be on the order of 1067 years! (The current age of the universe is only 1.4 x 1010 years.)

**Evidence for the existence of black holes**

Given the blackness of black holes, how would one go about looking for them? Hawking radiation will be much too weak to observe, so it seems that there will be no opportunities to observe the black hole directly. But there is enough less direct evidence to convince scientists that black holes really do exist.

If matter falls into a black hole, it likely will spiral in toward the event horizon. This spiraling gas will emit lots of X-rays. Initially the search for black holes focused on visible stars with binary partners that gave off lots of X-rays but no visible light. Several candidates emerged, and most of those are now believed to be black holes. They have sufficient mass and are very bright in X-rays. Future work involves a new space-based X-ray telescope that will have enough resolution to see the structure of the disk of matter accreting around the black hole. An accretion disk around a neutron star should show a bright X-ray source at the center as the infalling matter collides with the surface of the neutron star. An accretion disk around a black hole will be dark at the center.

The very large black holes at the center of most (if not all) galaxies have millions to billions of times the mass of our Sun. There are several lines of evidence pointing to their nature. For one, they are very compact. At the center of the Milky Way, for instance, is an object with a mass of about 2.5 million solar masses. We can observe stars nearby orbiting it with periods of 5 to 10 years, so we feel confident about the mass estimate. Those orbiting stars come very close to the central object without hitting anything, so it’s got to be pretty small. In addition, this object gives off lots of radio waves. We can see them because they can penetrate the dust that lies in the plane of the galaxy and blocks our view in visible wavelengths. Plus, for one “supermassive” black hole, a theoretical model of the X-ray spectrum has been constructed that includes a number of relativistic effects, and it fits amazingly well. So not only do we not have any other idea about what could be at the centers of galaxies to produce the radiation, we also have quite a detailed spectral match. No one believes this could be just a coincidence anymore.

Finally, there are objects in the sky called quasars (mentioned in an earlier section) that are enormously far away, and to be seen they must be converting mass to light at an incredible rate. When people tried to calculate the rate of nuclear reactions that would be needed to produce this sort of energy, the rate was too huge to be plausible. (Nuclear reactions generally convert about 1% of their mass into other forms of energy.) Only the types of gravitational interactions near a black hole can produce electromagnetic energy at the rates observed with plausible amounts of matter falling into a black hole. But these black holes likely have one *billion* times the mass of our Sun. As much as 30% of the mass falling into such a black hole can be converted into electromagnetic energy. (That infalling matter gets heated up so much that it outshines its entire surrounding galaxy. That’s why a quasar appears to be a point-like object in images: the rest of the galaxy doesn’t show up because the exposures are just too short.)

All of these lines of evidence (and some more not mentioned here) have convinced people that black holes are real. Stephen Hawking made a famous bet with Kip Thorne back in 1974, soon after Cygnus X-1 was discovered, about the nature of that object. Thorne claimed it would be found to be a black hole, while Hawking thought otherwise. In 1990, Hawking, with some assistance, broke into Thorne’s office at Caltech and wrote a concession on the original written copy of the bet. Since then, the evidence has only gotten stronger.**Exercises**

6.1 What is the Schwarzschild radius for a black hole with the mass of the Sun (2 x 1030 kg)? Of the Earth (6 x 1024 kg)? Of the black hole at the center of our galaxy (about 2.6 million solar masses)?

6.2 Suppose a source of light is located at a distance r = 3GM/c2 from a black hole. The frequency of the source is 1 x 1015 Hz, just into the ultraviolet so that it normally cannot be seen with the eye. What frequency will be observed by an observer at infinity? Will you see the light? If so, what color will you see?

6.3 Suppose a source of light is located at a distance r = 2GM/c2 from the center of a black hole. The frequency of the source is 1 x 1015 Hz, just into the ultraviolet so that it normally cannot be seen with the eye. What frequency will be observed by an observer at infinity? What is the significance of this answer?

6.4 In order to jump into the future and escape the statute of limitations (it wasn’t anything serious, really, just a little quantum mechanical practical joke), Phred the physics student travels to a black hole. He gets close enough to the black hole so that he will experience just one month for every year that goes by at home. If the mass of the black hole is 8 x 1030 kg, at what radius will he need to be? (You may assume he will just hover there, so that his velocity is zero.) You may answer either in meters or in multiples of the Schwarzschild radius.

6.5 Use the time dilation formula and a Maclaurin series to show that the dilation factor is just the square root of the altitude above the event horizon divided by the Schwarzschild radius, as long as the altitude is small compared to the Schwarzschild radius. (Hint: Make the radius = rSchwarzschild·), where  is very small compared to 1.)

6.6 What is the rate of evaporation (mass lost per second) for a stellar sized black hole? For the black hole at the center of our galaxy (about 3 million solar masses)?

6.7 Since a black hole evaporates faster and faster, you cannot use the current rate of evaporation to directly find the lifetime of a black hole. Can you use a spreadsheet (or iteration on a calculator) to estimate the lifetime of a solar mass black hole?

6.8 Use calculus to derive the formula for the lifetime of a black hole that depends on the mass of the black hole. Find the lifetime of a solar mass black hole.

6.9 Here’s another weird thing about black holes: Black hole thermodynamics!   
  
a) If you add energy to a black hole, what happens to its mass?   
  
b) What happens to the rate at which it emits Hawking radiation as a result?   
  
c) So what has happened to the temperature of the black hole?   
  
d) Why does this seem weird?

Chapter 7. Cosmology

Gravity is the force that explains the motion of bodies on the largest scales. The nuclear forces are too short ranged, and although the electromagnetic interaction is a long-range force, its strength makes it unlikely for large bodies to accumulate very much net charge. The electrostatic attraction between a proton and an electron is greater by about 40 orders of magnitude than their attraction due to gravity. Charges of one sign quickly attract charges of the opposite sign and so negate each other over astronomical distances. Gravity, however, cannot be shielded – there is no negative mass – so gravity’s influence is felt out to the farthest reaches and controls the fate of the universe.

**The expanding universe and the cosmological constant**

When Einstein finished his general theory of relativity he knew that his theory had a chance to help us understand the features we observe in the cosmos. But at that time it was not known that the universe was expanding. People believed in a “steady state” model in which the universe had essentially always been the same. Historically, this was considered a progressive view in that there was no need for a moment of creation. (It was also unclear how one would handle the apparent discontinuity in time that a moment of creation would imply.) When Einstein applied his field equation to a universe of uniform density, the result was unexpected. Here is the formula he derived:

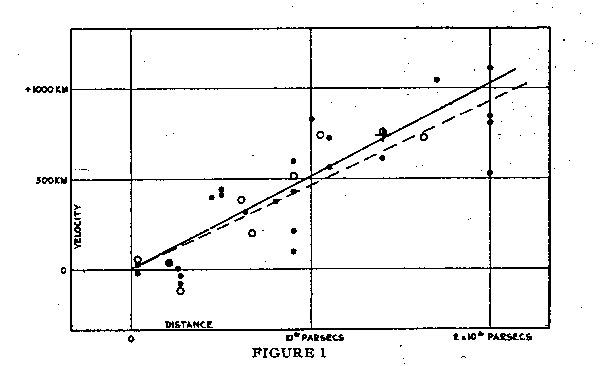
.

The variable ‘R’ represents the distance between two points in space;  is the average density of the universe. The left side of the equation is the square of the rate of change of R. (For those who haven’t encountered the notation of calculus, dR/dt is roughly the same as R/t, the change in R divided by the time it took for the change.) In a static universe this would be zero. But the right side is zero only if the average density of the universe is zero! Since we exist, we will assume that the average density is greater than zero, and so a static universe is not possible. R has to change; the universe cannot be static. An interesting feature is that the rate of change of R can be either positive or negative – it gets squared, so the result is positive either way. According to this result, the universe must either be expanding or contracting. And here is where Einstein broke from his normal way of doing things.

In all of Einstein’s previous work, whatever results he obtained from his postulates were what he lived with. When a student one time asked him how he would feel if the bending of light were not upheld by experiment, he replied, “I would feel very sorry for God; the theory is right.” It didn’t matter if the results seemed counterintuitive. They were the results, and you had to live with them.

In this case Einstein took a different course. He found a way to insert into his field equation a ‘fudge factor’ , which he called the cosmological constant. It provided a repulsive effect to balance the attraction of gravity at large distances, allowing for the universe to be static. Mathematically it provided a negative term on the right to allow dR/dt to be zero. Physically it represents an energy associated with empty space that provides an outward pressure proportional to the volume of free space. In this way, the effect of the cosmological constant would not appear at small separations. Only when one looks at the universe on the largest scale would its effect appear. By choosing an appropriate value for  the galaxies are held at bay by balancing their gravitational attraction with the outward push of the cosmological constant. In this way Einstein avoided the problem of an initial time, a moment of creation which his mathematics couldn’t explain. So the cosmological constant was introduced in order to obtain a desired result – there was no *a priori* philosophical reason for inserting this term. And the same goes for choosing its value: you just had to fiddle with it until you got a stable universe of the right size.

About 12 years after Einstein published this revised formula Edwin Hubble, working at the Mt. Palomar observatory in California, uncovered evidence that the universe was indeed expanding. Hubble observed that more distant galaxies were moving away from us at a faster speed as determined by the redshift of their spectral lines.



From the Proceedings of the National Academy of Sciences  
Volume 15 : March 15, 1929 : Number 3

Notice that there is a general linear appearance to the data. The slope of this line is called the Hubble constant, and although Hubble’s data suggested a value near 500 km/s per megaparsec, its currently accepted value is in the neighborhood of 70 km/s per megaparsec. Why the difference? The motion of local galaxies is at such a speed that it masks the cosmological expansion. To see more clearly what the rate of expansion is you need to look at more distant galaxies that Hubble couldn’t see with his telescope. The local motion can be large enough to completely overcome this expansion, as with the Andromeda galaxy, which shows a negative relative velocity. Our galaxy and Andromeda are approaching each other and will collide in a few billion years. (Note: Although we call it the Hubble constant, it is not necessarily a constant over time. In the model prior to the cosmological constant, the expansion of the universe is slowed by gravity and the Hubble constant decreases with time.)

Why does this relationship imply the universe is expanding? In the equation dR/dt is proportional to R – the farther away two points are, the faster they will be moving from each other. Imagine putting adhesive dots randomly on the surface of a balloon, and then inflating the balloon. If the balloon’s diameter increases by 10% over some interval, then all the distances between galaxies will increase by the same 10%. This means that more distant galaxies will have moved a greater distance in the same time: the velocity of recession should be proportional to distance, just as Hubble observed through his telescope. Hubble’s evidence of the expanding universe has been validated many times over since his discovery. The universe on the whole is expanding.

Einstein’s original equation had given him the correct answer, and he had refused to accept it. Upon hearing of Hubble’s results he called the cosmological constant “the biggest blunder of my life.” The cosmological constant was quickly thrown out. (In addition, the cosmological constant was later shown to provide only an unstable equilibrium. This made it unacceptable since it is unlikely that any system, much less the entire universe, would have remained in an unstable equilibrium for long.)

**Big Bang**

What does this tell us about the universe? If it is expanding now, it must have been smaller in the past. And since things cool as they expand, it must have been hotter in the past than it is now. If you take Einstein’s equation and use it to model the evolution of the universe, you find you can only go back so far in time – at some early time the universe would have had a size of zero! So a simple reading of Einstein suggests that there was some initial time at which the universe began. The density and temperature were infinite. The universe began in a state of rapid expansion. These infinities and the idea of an earliest time struck people as very strange. (Infinities like these suggest that a theory isn’t complete.) One astronomer who believed in a steady state universe, one that always looked like it looks today, was Fred Hoyle. In a radio interview he derided this crazy idea that there was some “big bang” that started the universe. The name caught on. The Big Bang is now the accepted story of how the universe has evolved over the past 14 billion years or so.

So if the universe began as a point (or at least very compact and highly dense), *where* did it begin? Here is a wild idea that comes straight out of Einstein: the expansion of the universe we have discussed is not the type of expansion you get if a bomb explodes. In that situation the fragments of the bomb fly away through a preexisting space. Space, in that case, is the stage upon which events occur. But in Einstein’s world, space itself is expanding, carrying galaxies along with it like a conveyor belt. Go back to your imaginary balloon (or a real one if you’d like). Imagine a video of the inflating balloon played in reverse. For a mathematically perfect balloon, its radius and surface area get smaller and smaller until at t = 0 the balloon (universe) becomes a point. Every place in the universe came from that point; every point in the universe is equally the center of the expansion; every point is equally the center of the universe.

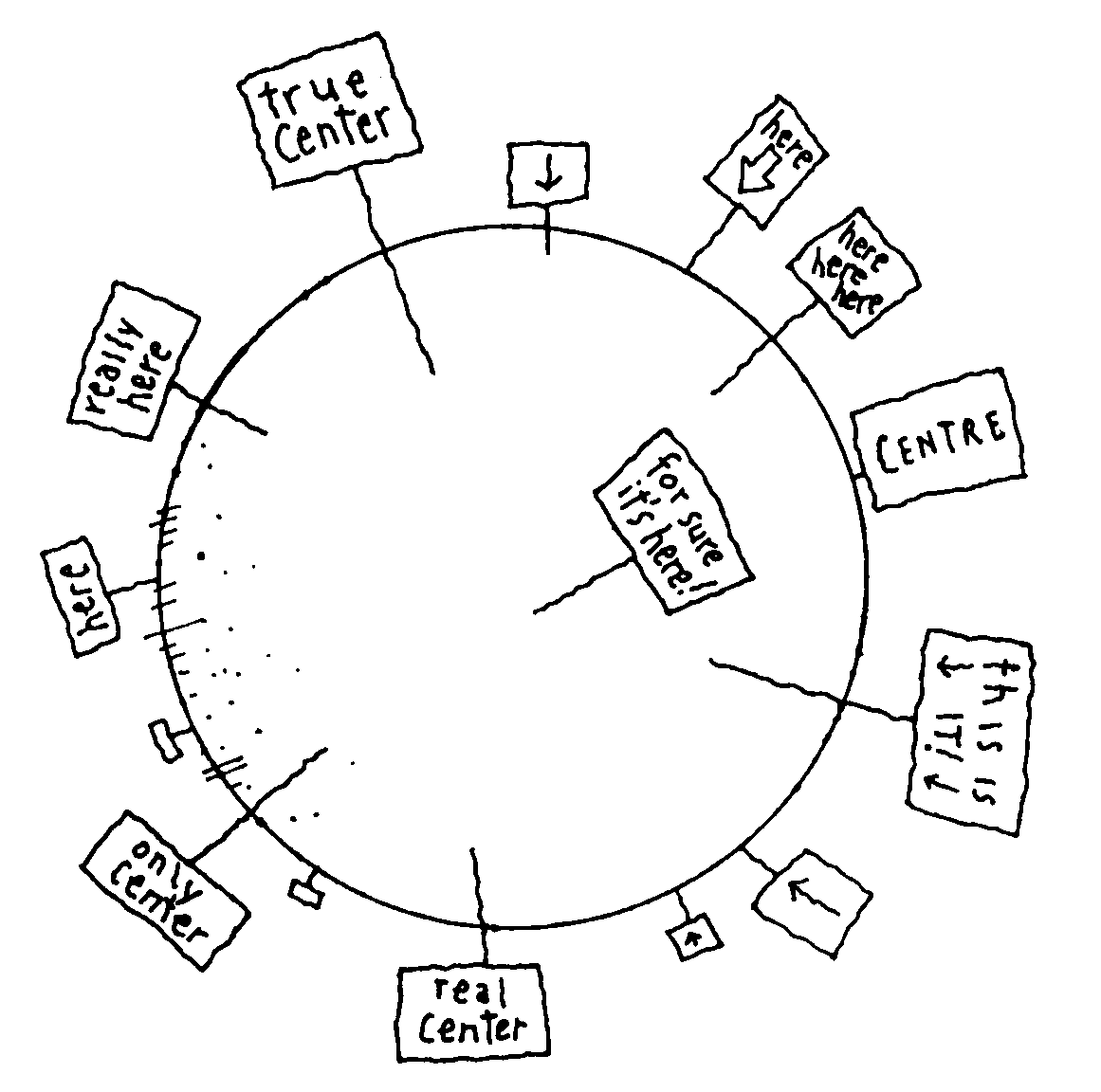


Figure taken from Rudy Rucker, The Fourth Dimension, 1984. (Permission not yet requested.)

In this analogy, the two dimensional surface of the balloon clearly expands into three dimensional space. The natural question arises: What is the universe expanding into? Well, it may be expanding into some higher dimensional space, but it doesn’t have to. (This brings up a problem with analogies. They are designed to get one point across, and they run the risk of communicating incorrect ideas as well.)

Here’s another model of the expansion of the universe. Start with a set of dots randomly scattered in a square. As before, these dots represent random galaxies in the universe. (In the images below, these are open circles.) Now take the image of the square and make a copy that is enlarged by 10%. (These are the black circles below.) This represents the universe after some time has gone by.

Original locations of galaxies

Locations of galaxies after 10% expansion

Overlay the images so that one galaxy, call it A, is lined up in the two images. You can see that all the other galaxies now have double images. The separation of the two images of each galaxy show how far that galaxy has moved relative to the home galaxy in the amount of time that has passed between the two images. You can see that the more distant galaxies moved farther than the closer ones; the speed of recession is proportional to the distance. This is the Hubble relationship again! If you try this again, say with galaxy B, you can see that it wasn’t just luck that I happened to pick the right one. Observers in *any* galaxy see all the other galaxies moving away from them as time goes by; all are at the ‘center’ of expansion.

A

B

A: I’m at the center! B: No, I’m at the center!

Here’s the amazing part: Each set of dots shows just a part of a two-dimensional universe. Imagine the universe to fill an infinite plane. When you expand the universe by 10%, how much bigger is it? It is still infinite. It doesn’t actually take up any more space! It still exists in the same two dimensional plane. (There are different sizes of infinities, but expanding an infinity by 10% doesn’t make it a bigger infinity.) This means that a universe can expand without having to expand into any new, preexisting territory. It can simply expand. This is really weird, but it works. You just have to get used to how big infinity is.

So the space of the universe, the three dimensions in which we live, may have come into existence at the moment of the big bang. Space did not precede the universe (as Newton might have presumed), but was created at the same time all the matter of the universe came into being. (Quantum mechanics is likely to modify this last statement.)

**A really weird implication**

Imagine a rubber rope with one end tied to a post. The other end is hooked to the bumper of a truck that is accelerating away. The rule for expansion of the universe, the Hubble Law, is that the average speed, ignoring local motions, is proportional to the distance: vspot on rope = Hx, where x is the distance from the observer (the post in our case) and H is the rubber rope value of the Hubble constant. What does this say about the speed of some spot on the rope for very large distances? It would seem that for a sufficiently distant point, the velocity will be greater than the speed of light.

Seems like a contradiction with Special Relativity, doesn’t it? Fortunately for General Relativity, we get off on a technicality. In Special Relativity, the rule is that nothing can be observed moving relative to you faster than the speed of light. If there is a galaxy sufficiently far away that it is moving away faster than light, you will never observe it moving that fast because light from it can never reach you. Saved! As a result, we don’t talk about the size of the universe any more. We very carefully speak only of the size of the observable universe.

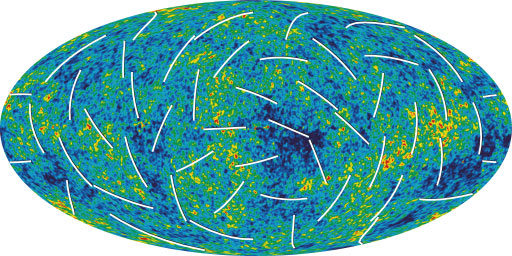
**Time**

Okay, so trying to imagine what existed if space didn’t exist is tough. But thinking back to Special Relativity, we learned that space and time are not separate things. Guess what? The big bang was not only the beginning of space – it was the beginning of time, too.

How can this be? One problem with thinking about time having a beginning is the thought that time would be discontinuous. There would be an after, but no before. Stephen Hawking has an interesting way of thinking about this. Start by picturing the globe of the Earth. From just about anywhere on the globe you can head north. How far north can you go? All the way to the North Pole. And that’s the farthest north you can go. If you continue traveling in a straight line, you wind up going south instead. There is a northernmost point on the globe even though the surface of the globe is continuous. Hawking imagines the moment of creation to be analogous to the North Pole. There can be an earliest time, and proceeding from that earliest time will necessarily lead you into the future. There is no time earlier, and yet time is not discontinuous. (It’s okay if you don’t like this idea; not everyone does, but it is a possibility.)

What got it started? No one knows. (One unsatisfying response to this question is that the question itself is faulty – a cause comes before, and there was no before!) This question, like what happens at the center of a black hole, looks at a situation where both quantum effects and strong spacetime curvature are involved. We really don’t know what the answer is. (See a later lecture for conjectures on this.) In addition, Einstein’s formulas really only tell us how the universe ought to evolve once the initial conditions are in place. As noted earlier, there are lots of infinities associated with t = 0, so we really don’t believe the results that far back. We don’t know if t = 0 was really the beginning of time and space.

In spite of this ambiguity, Einstein’s equation does a very good job of modeling the expansion of the universe, and the basic “hot big bang” picture is well established, beginning a fraction of a second after the big bang. An additional piece of evidence that backs up the big bang model of cosmology is called the cosmic microwave background radiation (CMB). It is a remnant from the time when the early universe first became transparent to electromagnetic radiation. Initially the wavelength of the CMB was much shorter than it is now, but the stretching of space has increased the distance between the crests of the electromagnetic waves (thereby dropping the temperature) of this ancient light. The temperature of the CMB is now about 2.7 Kelvin, which fits with calculations made prior its discovery by Penzias and Wilson at Bell Labs. The temperature variation of the CMBacross the sky, as observed by more recent and precise experiments, is only 1 part in 100,000, suggesting that the early universe was incredibly homogeneous. The map of this radiation looks mottled, with spots of very slightly warmer (red) and cooler (blue) regions scattered across the sky.



Credit: NASA/WMAP Science Team

**Overall curvature of the universe**

In the early days of applying General Relativity to the universe as a whole, it was realized that there could be a curvature to the universe. Further, this curvature would be important in regard to the fate of the universe. If the average density of matter in the universe (designated ) is high, the overall curvature will be positive and the universe will slow down in its expansion and recollapse. If the density of matter is low, the curvature will be negative and the universe will never stop expanding. If the density is just right (the critical density, c, about 5 H atoms per cubic meter on average today), the universe is like a projectile with escape velocity: it will slow down and come to a stop after an infinite amount of time. (We’ll be modifying these statements in just a bit, so don’t place your complete trust in them. This is just the result from the simple form of Einstein’s equation.)

So what is the evidence for the density of matter? Is it enough to close the universe and make it recollapse? One way to tell would be to create a triangle 5 billion light years on a side and measure the angles at the vertices. Their sum will tell us about the curvature of the universe on a large scale. But that’s not very practical. So what can we do?

Going back to what we know of non-Euclidean geometry, the formula  works well in Euclidean space. If we just knew the size of something a long, long way away we could measure  and see if it is too small, too big, or (a la Goldilocks) just right. Well, we’ve got such a thing: the variations in the cosmic microwave background radiation. By knowing their size and how far away (how long ago) the CMB was emitted, we can plug s and r into the formula and solve for . The result from this calculation, about 1º, is in agreement with the WMAP data and is consistent with the universe being Euclidean. You can call it flat as well, but remember that it is flat in the sense that the geometry is Euclidean. The universe is not flat in the way that the Milky Way galaxy is a flat disk.

**What is the matter?**

Okay, the universe is flat in the Euclidean sense, or at least incredibly close to it. That means the density of the universe must be close to the critical density, c. If we try to add up the mass we can see in the universe, how close do we come to this value? Well, we run into trouble here. Beginning with the Milky Way, our home galaxy, when we add up the masses of the stars, clouds, and suspected black holes, we come up with a rather small mass compared to expectations. But there is another method of finding the mass of a galaxy, and it comes from Newton. By observing something orbiting the galaxy you can find the mass of the galaxy, just as you found the masses of planets from their satellites’ orbits in the study of planetary motion. There are globular clusters that orbit all galaxies that contain thousands to millions of stars. This makes them bright enough to see at great distances. By finding their speed and distance from the center, you can find the mass of the galaxy. The mass that comes out of that analysis is much greater than the mass you get by adding up the component parts. So we apparently missed a few things, but the galaxy is a crowded place and it is hard to see everything.

But there’s a twist: If we look at several globular clusters at increasing distances out from the galaxy, we get larger and larger estimates for the mass of the galaxy, even though (to the telescope) the space beyond the edge of the galaxy is empty. This implies that the empty space isn’t so empty. There must be some matter we can’t see out there, and it has been dubbed “dark matter” because it doesn’t emit any light. We’ve looked at many other galaxies and seen the same sort of result.

It doesn’t end there, though. When we look at clusters of galaxies, we find that the galaxies are moving so fast that the individual galaxies should have flown away from each other long ago. (Their speed exceeds the calculated escape velocity for the cluster.) The only way for these clusters to still exist is if there is more mass than can be accounted for even with the inflated values for the galaxy masses found by the globular cluster method. So there seems to be a lot of dark matter (more than in the galaxies themselves) filling in the space between the galaxies, and this dark matter provides enough gravity (or local curvature) to hold the cluster of galaxies together.

There is more evidence: We have observed clouds x-ray emitting gas between the galaxies in a cluster. To emit x-rays, the gas must be very hot (the molecules have a high average kinetic energy). In order to keep the clouds as dense as they are, there must be sufficient gravity present. The inferred mass of the cluster is in agreement with the earlier estimate based on escape velocities.

And more evidence: By looking at gravitational lensing as light passes through clusters of galaxies we see that the amount of bending is consistent with the amount of mass inferred from the presumed stability of those clusters.

So there are multiple lines of evidence (galaxy masses, hot intergalactic gas, and gravitational lensing) for the existence of dark matter, and all support the idea that dark matter makes up about 5 to 7 times more mass than visible matter. In other words, most of the matter of the universe has never been seen! And there is evidence from the ratio of H to He that resulted from the big bang that this dark matter is not even made up of the usual protons, neutrons, electrons, and neutrinos. No one knows what it is, but there are a few suggestions that are being looked into. The amazing thing is that this dark matter, whatever it is, tends to cluster in the same way that galaxies do. That means the room you’re sitting in (assuming you’re reading this from inside of a galaxy) probably has dark matter in it! A remaining problem is that even with this amount of dark matter, the total matter density in the universe is only about 30% of the critical density. Since all of the lines of evidence point to the same amount of dark matter, we do not expect to find more. We will see where the remainder of the critical density may come from in the next sections.

**Accelerating expansion**

So now we’ve got a model for how the universe evolved over billions of years. How many billions? Astrophysicists again began with the simplest model. They took the current rate of expansion, the existence of dark matter, and worked backward. Since gravity is an attractive force it will slow the expansion; the universe would have been expanding faster in the past. Mathematically one can use Einstein’s equation to calculate how long ago the universe was of zero size. The result is in the area of 10 to 12 billion years ago.

Other researchers took a different approach to finding the age of the universe. They were studying globular clusters. These groups of stars are some of the oldest structures in the universe. By looking at the population of stars of various masses, they were able to determine that these clusters were, in some cases, around 14 or 15 billion years old.

That’s a problem. The constituents of the universe should not be older than the universe itself. There was some uncertainty in each of the numbers, but it didn’t seem that the amount of uncertainty was enough for these two figures to agree. People were hoping for some reason to believe that the universe was older than the age found using Einstein’s equation.

The thought occurred that perhaps the universe didn’t expand quite the way the simple theory suggested. What was needed was a reliable way to gauge the distance to the furthest galaxies to see what their redshift would say about expansion in the past. The tool that was found was the Type Ia supernova. There is more than one kind of supernova, but the Type Ia supernovas are believed to be of the pretty much the same brightness each time. (They occur by the gradual addition of mass onto a neutron star. When it gains enough mass, it implodes into a black hole. It’s the same mass each time, so the explosion is of the same approximate strength each time.)

The most distant of these supernovae had an interesting characteristic. For their redshift, you compute a distance from the simple model of expansion. And from the distance, you can then predict what the apparent brightness should be. But some of these distant supernovae were too faint for their redshift. The faintness said that they must be further away than simple expansion would predict. If we are speeding up now after a period of slowing, then the universe could be older and bigger than the simple model predicted. (The data is fairly good on this: there have been enough of these special supernovae observed to map out a rough curve for expansion that shows an initial slowdown, as expected, followed by a speedup beginning about 5 billion years ago. More precise data would help test our models better. The Hubble telescope can help with this if it remains in operation for a few more years.) But if gravity is always attractive, its only role should be to slow things down. What could be causing the universe to accelerate?

**According to an actual fortune cookie I got: “Matter” is not what matters most**

If the universe is Euclidean, then omega = 1. If the density of matter (dark and normal combined is 0.3, then something else must make up the rest. An energy spread throughout space would work with matter to determine the curvature (or lack) of space, but it would work against the attraction of gravity. Since we can’t see it we call it *dark energy*. We don’t know what it might be. Oddly enough, the effect of Einstein’s cosmological constant produces a model of the universe that matches the supernova observations. So perhaps Einstein’s “biggest blunder” will turn out to be right after all.

What value of  would we like? We would want it to be big enough to account for the missing 70% of the critical density (about 3.5 Hydrogen atoms per cubic meter, but not in the form of Hydrogen atoms). Since this dark energy is uniformly distributed through space, it would not cause bending of light or distortion of orbits. But it would contribute to the overall curvature of space. So dark energy would resolve the question of what constitutes the missing piece of cosmic density. In addition, the value that works for the density also affects the expansion of the universe in just the right way to match the dimness of the distant Type Ia supernovae. And as a bonus, it also allows the equations to accurately model the geographic (or perhaps we should call it cosmographic) distribution of galaxies. So there are three separate puzzles that the existence of dark energy would solve.

But even if the cosmological constant gives us a model that matches observations, it is still a problem in that we’re introducing it only to fit what we see. There is no *a priori* justification for it. Quantum physicists have made a guess at why there should be such a term. They think it comes from “vacuum energy” which is due to quantum effects. When early attempts were made to calculate just how big the term should be, they were off by 120 orders of magnitude. They did get the sign right, but that’s about all. (To give you an idea of just how far off they were, the size of a proton and the size of the entire observable universe are different by about 41 orders of magnitude. They were off by that ratio about three times over.)

Bottom line: The overall content of the universe seems to be about 73% dark energy, 23% dark matter, and only 4% matter of the type we understand: protons, neutrons, electrons, neutrinos. The age of the universe is 13.7 ± 0.2 billion years.

What are the consequences of an accelerating universe? Think back to the rubber rope analogy. If the Hubble constant were to stay the same over timed, then as a galaxy recedes from us it would be moving faster and faster. Eventually galaxies that used to be observable would be moving away from us at a speed that exceeds c. And we would no longer be able to see them. If what we are observing is really due to a cosmological constant, then astronomers 50 to 100 billion years in the future would only see a handful of the hundreds of billions of galaxies we observe today.

**Inflation**

There are two other issues that we have glossed over so far. One is called the ‘horizon problem.’ When we look at the CMB in one direction, we see blackbody radiation at 2.73 K. When we look in the opposite direction, we see the very same thing. But these two regions are so far apart that light from one region would never make it to the other region given the age of the universe. If light couldn’t, neither could any other effect. It is said that these regions are not *causally connected*. Normally two objects reach equilibrium by being in contact. So how is it that these distant regions seem to be in equilibrium?

The other issue is the flatness of the universe. If the density of the universe is just right, then we get a flat universe. But as the universe expands, the density changes. A flat universe is in an unstable condition. If your universe starts out just a little bit denser than critical, it will evolve away from flatness toward a positively curved universe. If the density is initially just below critical, the curvature will become more and more negative over time. So here we are, 13.7 billion years after the big bang, and the universe is still very nearly flat. That either requires some amazingly good luck on our part, or else something acted to produce this result. (Imagine a marble rolling down a knife edge for that long without falling to either side. Would you simply say it was lucky, or would you look for a cause?)

A possible answer that seems reasonable at the present time is called *inflation*. It solves the horizon and flatness problems by postulating an early exponentially accelerated expansion of the universe. In this way, the size of the universe would have grown by a factor of 20 to 40 orders of magnitude in a matter of 10-34 seconds. This would have been driven by a term that acted like a huge cosmological constant, but which turned off after this very short burst of expansion. (This expansion of space far outstrips the travel time at the speed of light and would separate regions that had been in contact with each other.) The universe would then have reverted to an expansion that follows a decelerating pattern for several billion years. Our observable universe would be just a small region of what resulted from the accelerated expansion. Regions of the universe that are not causally connected today would once have been close enough initially to interact, allowing them to reach equilibrium. It is as though O’Malley and Ophelia, talking to each other before the universe reached the age of 10-36 seconds, were suddenly whisked away from each other by the expansion of space. By 10-34 seconds the accelerated expansion was over, and it was as though their friend had disappeared. Then, perhaps a billion years later, light signals from each of them were finally able to catch up to the other, and they could exchange messages again. (But the expansion is accelerating again now, and they will eventually be lost to each other again.) What would effect would the expansion have on the curvature of the universe? When you take a balloon and inflate it, the curve of its surface becomes flatter and flatter. The very early exponential expansion would have produced a universe that is very nearly perfectly flat, no matter what the starting conditions.

And there are still other issues that the inflationary model helps with, too: 1) Many models of a unified theory (one that unites general relativity with quantum mechanics) require the existence of magnetic monopoles at a density that should make them commonplace, and yet they have never been observed. Inflation would have spread out the monopoles to a density of less than one monopole per galaxy, making the lack of observations less of a problem. 2) Where did all the mass of the universe come from? The odd behavior of the quantum field that would cause inflation would help explain how a small amount of mass initially (even just 10 kg) could be turned into all of the mass in the entire observable universe. 3) If the inflationary expansion smooths everything out, how did the variations in density that allowed for the formations of galaxy clusters come about? Normally quantum fluctuations are microscopic in size, but when the rapid expansion of inflation occurred these fluctuations were blown up to literally astronomical size. The large scale structure of the universe seems to have come from greatly enlarged quantum jitters. 4) We have always wondered how the universe could have started in such a low-entropy state. Since gravity causes lumps to form, a smooth universe is unlikely and hence is a low-entropy state as far as gravitational entropy is concerned. Inflation gives us that low-entropy start.

The idea of inflation is now part of most models of the evolution of the universe. It still doesn’t take us back all the way to t = 0, but it works nicely as a ‘front end’ that provides the conditions from which the big bang proceeded.

**Topology**

Topology is the area of geometry that investigates connections. In the realm of cosmology, the question relates to the overall connections: is the universe infinite or finite? If it is finite, it need not have a boundary. Remember the case of the sphere, where you can travel forever without ever finding an edge. But what would you find? You’d find that you return to the same place after a while. Could this be true for the universe?

It would need to be a pretty big sphere. One source of evidence for or against this may come from a careful analysis of the CMB. There would be patterns to this radiation if the universe is in the right size range. So far those patterns have not been seen in the data from the Wilkinson Microwave Anisotropy Probe (WMAP). There will be more to come on this in the next decade as new and more sensitive space-based telescopes are launched.

**References**

Harrison, Edward. Cosmology: the science of the universe, 2000.

Levin, Janna. How the Universe Got its Spots, 2002.

**Exercises**

7.1 What is it like at the edge of the observable universe?

7.2 If everything in the universe were expanding at the same rate, would we know it?

7.3 What would the universe look like if instead of having space expand, atoms were shrinking?

7.4 What part(s) of the universe is (are) expanding?

7.5 Let’s accept that the universe is 13.7 billion years old. If telescopes could see to a distance of 13.7 billion light years, what would they show?

7.6 Suppose that the Hubble constant were to remain constant over time. (This would be true in a universe dominated by a cosmological constant.)   
  
a) Using the formula given for the rubber rope, vspot on rope = Hx, show that the   
 distance to another galaxy will grow exponentially with time. (Hint: You will   
 need to use calculus.)   
  
b) What does this say about the recession velocity of a given galaxy as time goes   
 by?

7.7 Assume that the force due to the cosmological constant grows with distance. Imagine that at some distance D the force of gravity and the outward cosmological force balance. Why is this equilibrium unstable?

7.8 Using Einstein’s equation, show how the Hubble constant depends on density.